UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Predicción en modelos de componentes inobservables condicionalmente heteroscedásticos

Autor:
Santiago Pellegrini

Director/es:
Esther Ruiz Ortega
Antoni Espasa Terrades

DEPARTAMENTO de ESTADÍSTICA

Getafe, Abril de 2009

To my family and my wife Belén

Contents

List of Figures

List of Tables

Acknowledgements

This thesis would not be finished in time and shape without the help and support of several people whom I would like to thank and to whom I am happily in debt.

First of all, I would like to thank my advisors, Dr. Esther Ruiz and Dr. Antoni Espasa for supporting, guiding, and working very closely with me for the last four years. I am very grateful for their patience, encouragement, criticism, persistence, and friendly relationship with me. They will always be very good friends to me.

I am very grateful to the entire faculty and staff of the Department of Statistics for their support in every step of the way, for always being friendly, helpful, and appreciative of our work. A special mention to all my colleagues and friends that shared with me many of the good times I spent here during my thesis work, and offered me support and advice. I could never forget our ability of converting simple coffee breaks into passionate debates on a wide range of topics, including politics, sports, technology, movies, and music.

I would also like to show my gratitude to the Spanish Ministry of Education and Science and to the education and culture section of the Community of Madrid for financial support. In particular, I would like to thank Professors Esther Ruiz and Antoni Espasa for allowing me to participate in the research projects SEJ 2006-03919 and CAM 2006-03560.

I dedicate my thesis to my family and specially to my parents, Daniel and Alicia, who have always supported me in all my decisions, even in the one that implied coming to Madrid to go on with my graduate studies, far away from them.

My most special dedication goes to my dear wife Belén, who has accompanied me since the first minute I decided to come to Spain. Without her love, support and patience probably I would have never finished my PhD.

Chapter 1

Introduction

During the last two decades, there has been an increasing interest in the academic and practitioner world on modelling the volatility clustering observed in many economic and financial series. This research was pioneered by Engle (1982) and Bollerslev (1986), with the introduction of GARCH models. It is also common to observe stochastic trends in many economic and financial time series. In this case, a popular practice is to take differences in order to obtain a stationary transformation. Then, an ARMA model is fitted to this transformation to represent the transitory dependence. Alternatively, the dynamic properties of series with stochastic trends may be represented by unobserved component models. It is well known that both models are equivalent when the disturbances are Gaussian. In this case, the reduced form of an unobserved component model is an ARIMA model with restrictions on the parameters; see, for example, Harvey (1989). The main difference between both specifications is that while the ARIMA model includes only one disturbance, the corresponding unobserved component model incorporates several disturbances. Consequently, working with the ARIMA specification is usually simpler. However, using the unobserved components model may lead to discover features of the series that are not apparent in the reduced form model because they arise when estimating the components.

When combining both, stochastic trends and volatility clustering, the ARIMA and unobserved component models are not in general Gaussian. This implies that they are no longer equivalent when allowing for conditional heteroscedasticity in the noises. Among the large number of works devoted to studying and applying models that combine these features, almost none of them made a comparative analysis between the two alternatives. This important issue remains somewhat unexplored. Therefore, we think that more effort should be placed in this respect, specially in what regards to forecasting performance. The study of this issue represents the main goal of this thesis.

From an empirical point of view, the presence of conditional heteroscedasticity in both, ARIMA and unobserved component models, has previously interested many authors. There is a large literature that considers ARIMA models with either GARCH or Autoregressive Stochastic Volatility (ARSV) processes in the disturbances; see Bollerslev et al. (1992), Bollerslev et al. (1994), Diebold and Lopez (1995), and Diebold (2004) for detailed surveys. On the other

hand, unobserved component models with GARCH disturbances have been receiving a lot of attention as they allow to distinguish which components are heteroscedastic. One of the earliest implementations of these models is Harvey et al. (1992), which consider latent factor models; see also King et al. (1994), Sentana and Fiorentini (2001), Chang and Kim (2004) and Sentana (2004) for other applications related with latent factor models. Chadha and Sarno (2002) and Moore and Schaller (2002) fit unobserved component models with GARCH disturbances to price volatility and term structure of interest rates, respectively. Additionally, unobserved component models with ARSV processes in one or more disturbances have also been studied and fitted to financial data; see for example, Koopman and Bos (2004) and Bos and Shephard (2006). With respect to inflation rates, the presence of heteroscedasticity in the transitory and/or permanent components is a broadly debated issue in the recent literature; see Broto and Ruiz (2009) for a detailed survey. For example, Stock and Watson (2007) find that a simple unobserved component model with conditionally heteroscedastic noises in the form of ARSV processes describe well the dynamics of the US inflation.

We will first focus on the study of the statistical properties of a general class of unobserved component model, the local linear trend model, which considers that the series of interest, y_t, is composed by a transitory component, ε_t, and a stochastic trend, μ_t, with a stochastic slope β_t, in the following way,

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, && \text{(1.1a)} \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t, && \text{(1.1b)} \\ \beta_t &= \beta_{t-1} + \xi_t, && \text{(1.1c)} \end{aligned}$$

where ε_t, η_t and ξ_t are mutually independent and serially uncorrelated processes, with zero means and variances σ_ε^2, σ_η^2 and σ_ξ^2, respectively. Throughout Chapters 2-4, we focus on two particular cases of model (1.1), which are of interest from an empirical point of view. The first one is obtained when $\sigma_\xi^2 = 0$. Under this set up, the slope is fixed and the trend reduces to a random walk with a drift given by β_0. If, without loss of generality, we also assume that $\beta_0 = 0$, then y_t follows a local level model

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, && \text{(1.2a)} \\ \mu_t &= \mu_{t-1} + \eta_t, && \text{(1.2b)} \end{aligned}$$

see, for example, Durbin and Koopman (2001) for a detailed description and applications of this model. In this case, taking first differences results in a stationary series given by

$$\Delta y_t = \eta_t + \Delta\varepsilon_t. \tag{1.3}$$

On the other hand, when $\sigma_\xi^2 > 0$ but $\sigma_\eta^2 = 0$, the smooth trend model is obtained

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, && \text{(1.4a)} \\ \mu_t &= \mu_{t-1} + \beta_{t-1}, && \text{(1.4b)} \\ \beta_t &= \beta_{t-1} + \xi_t, && \text{(1.4c)} \end{aligned}$$

see Harvey and Jaeger (1993) and Nyblom and Harvey (2001) for description and applications. In this case, the trend is an integrated random walk and therefore two differences are necessary to obtain the stationary transformation, that is

$$\Delta^2 y_t = \xi_{t-1} + \Delta^2 \varepsilon_t. \tag{1.5}$$

As our purpose is to study conditionally heteroscedastic models, we allow ε_t, η_t, and/or ξ_t to be either GARCH(1,1) or ARSV(1) processes. In Chapter 2 we derive the population moments of the stationary transformations given in (1.3) and (1.5). Given that conditional heteroscedastic processes are characterized by having a positive excess kurtosis and a significant autocorrelation function (acf) of squares, we focus on these two moments. Alternatively, we also derive the properties of the corresponding reduced form ARIMA disturbance, denoted by a_t, in terms of ε_t and η_t. We show that, in general, there is not a simple GARCH or ARSV model for a_t that replicates the statistical properties of the conditionally heteroscedastic local linear trend model. However, given that it is common to fit GARCH or ARSV processes to the ARIMA residuals whenever they show evidence of conditional heteroscedasticity, we carry out several Monte Carlo simulations and analyze at which level, these models are able to capture the dynamics of series with conditionally heteroscedastic unobserved components. The chapter also has an empirical illustration with a daily series of the Pound/Euro exchange rate.

Chapter 3 is devoted to study the comparative forecasting performance of the two alternative ways of modelling conditionally heteroscedastic series with stochastic trends. Using the properties of the local level and smooth trend models developed in Chapter 2, the objective of Chapter 3 is to compare the performance of prediction intervals obtained from these two unobserved component models with the prediction intervals constructed from the corresponding IMA models when the noises are GARCH. We first find the theoretical Mean Squared Forecast Error (MSFE) of the two competing models and then construct the prediction intervals k steps ahead, with k being the time horizon.

Because of the time-varying volatilities, the amplitudes of the intervals change depending on whether the conditional variance at the moment of forecasting is larger or smaller than the marginal variance. Denoting by excess volatility, the difference between both variances, in Chapter 3 we show that if only the transitory component is heteroscedastic, the excess volatility disappears as the prediction horizon increases. That is, the prediction intervals obtained with the unobserved component models converge to the intervals of the corresponding homoscedastic model. However, given that the reduced form IMA models always contain at least one unit root, the corresponding prediction intervals depend on the excess volatility for all prediction horizons. Consequently, when the excess volatility is positive (negative), the multi-step prediction intervals based on the IMA models are too wide (too narrow), in terms of the nominal coverage when compared with the intervals based on the corresponding unobserved component models. On the other hand, when the heteroscedasticity affects the long-run stochastic level, the prediction intervals constructed using both models are very similar. Chapter 3 also contains an empirical exercise with the Pound/Euro exchange rate and US inflation rate in order to illustrate the main findings.

As we mentioned above, it is by now rather popular to fit unobserved component models with

conditionally heteroscedastic noises to describe and predict the dynamic evolution of seasonally adjusted monthly prices; see, for example, Stock and Watson (2002, 2007) and Cecchetti et al. (2007) for recent references. These models are very flexible and lead to predictions of future prices with adequate properties. In line with the idea of approximating a nonlinear model by a linear one, with time-varying parameters, Stock and Watson (2007) propose an IMA model with time-varying parameters (TV-IMA) as an alternative to the conditionally heteroscedastic unobserved component model. Despite that the two approaches have been utilized in many empirical works, there is a lack of papers studying the implications of using a TV-IMA model to fit series with conditionally heteroscedastic unobserved components. In Chapter 4 we focus on this issue by comparing the conditionally heteroscedastic local level model with its reduced form IMA counterpart with time-varying parameters. In particular, we analyze the evolution of the reduced form MA parameter for different specifications of the local level noises. In particular, we consider the cases in which the local level noises are i) homoscedastic, ii) conditionally heteroscedastic but stationary, and have structural breaks in the marginal variances. We show that the evolution of the MA parameter, θ, detected by Stock and Watson (2007) can be only attributed to breaks and not to the evolution of the variances. However, given that the random walk model that they assume for the log-volatilities of the components is well designed to deal with structural breaks, they are able to explain changes in the MA parameters with the conditionally heteroscedastic local level model.

Chapter 4 contains an exhaustive out-of-sample forecasting evaluation of several models fitted to the monthly US inflation rate. Although these exercises have already been performed by many authors, our contribution is that we measure the forecasting performance of unobserved component and ARIMA models, considering not only the point but also the interval forecasts. In this regard, we believe that predicting the uncertainty surrounding a given point forecast is as important as the point forecast itself, specially for policy makers.

Another important topic on inflation forecasting that has been recently receiving lot of attention, specially in the Euro Area, is the effect of disaggregation on the forecasts of index series; see Espasa and Albacete (2004), Hubrich (2005) and Hendry and Hubrich (2006) for a detailed study of this topic and applications for the Harmonized Consumer Price Index of the Euro zone. Within the context of vector ARMA (VARMA) models, forecasting contemporaneous and temporal aggregations of different series has been widely analyzed since the seminal work of Lütkepohl (1987). On the other hand, aggregation in conditionally heteroscedastic models in the form of GARCH processes has also been considered in the literature. For example, Drost and Nijman (1993) and Nijman and Sentana (1996) analyzes the properties of GARCH processes under temporal and contemporaneous aggregation, respectively. In general, the results show that GARCH processes are not closed under these two types of aggregation. These results are also supported by Zaffaroni (2007), who analyze the effects of aggregation on a large number of GARCH processes under different assumptions about the relationship among these processes. On the other hand, Meddahi and Renault (2004) propose a new model based on linear state space modelling (or stochastic volatility modelling), that is closed under temporal aggregation. For an exhaustive review on temporal aggregation of VARMA, and GARCH models, see Silvestrini and Veredas (2008).

Within the literature on this topic, there are almost no attempts to study the effects of aggregating conditionally heteroscedastic noises on prediction intervals. In Chapter 5 we try to add some light on this topic by studying the forecasting performance of using a disaggregation approach when one or more components of the aggregate contain conditionally heteroscedastic noises. In particular, we analyze whether the disaggregation produces better prediction intervals than those constructed from the more common approach of fitting a model directly to the aggregate index. Although preliminary, the results on simulated series seem to support the use of a disaggregation approach in some particular cases because it yields more accurate prediction intervals than the approach of working directly with the aggregate.

Finally, Chapter 6 contains the main conclusions and some lines of future research that remain still open in what respect to prediction intervals for conditionally heteroscedastic unobserved component models.

Chapter 2

Statistical properties of conditionally heteroscedastic unobserved component models

2.1 Introduction

In this chapter, we are interested in analyzing the local level and smooth trend models given by (1.2) and (1.4) in the presence of conditionally heteroscedastic noises. In particular, we consider the cases in which either or both disturbances, ε_t and η_t, are GARCH(1,1) or ARSV(1) processes in the local level model, while in the smooth trend model we only allow for ε_t being GARCH(1,1). Given that conditionally heteroscedastic noises are characterized by having excess kurtosis and positive autocorrelations of squares, we derive these two moments for the stationary transformation of both models.

Consider the reduced form of the local level model given by (1.3). It is well known that this model has the same autocorrelation function (acf) as the following IMA(1,1) model

$$\Delta y_t = a_t + \theta a_{t-1}, \tag{2.1}$$

where, if Δy_t is invertible, then $\theta = \left[(q_\eta^2 + 4q_\eta)^{1/2} - 2 - q_\eta\right]/2$, with $q_\eta = \sigma_\eta^2/\sigma_\varepsilon^2$ being the signal-to-noise ratio. Note that the parameter θ is restricted to be negative, i.e. $-1 < \theta < 0$. Finally, the reduced form disturbance a_t is an uncorrelated process with zero mean and positive variance equal to $\sigma_a^2 = -\dfrac{\sigma_\varepsilon^2}{\theta}$.

On the other hand, the reduced form of the smooth trend model given by (1.5) has the same acf as the restricted IMA(2,2) given by

$$\Delta^2 y_t \;=\; a_t + \theta_1\, a_{t-1} + \theta_2\, a_{t-2}, \tag{2.2}$$

where the parameters θ_1, θ_2 and σ_a^2 are the solutions of the following system

$$\sigma_a^2(1+\theta_1^2+\theta_2^2) = \sigma_\varepsilon^2(6+q_\xi), \tag{2.3a}$$

$$\frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} = -\frac{4}{6+q_\xi}, \tag{2.3b}$$

$$\frac{\theta_2}{1+\theta_1^2+\theta_2^2} = \frac{1}{6+q_\xi}, \tag{2.3c}$$

with $q_\xi = \sigma_\xi^2/\sigma_\varepsilon^2$. There are four solutions of system (2.3) but only one of them contains a pair of real values for θ_1 and θ_2 that falls inside the invertibility region.

We also derive the properties of the reduced form IMA disturbance, a_t, in terms of ε_t, η_t and ξ_t. We show that, in general, there is not a simple GARCH or ARSV model for a_t that replicates exactly the statistical properties of the conditionally heteroscedastic unobserved component models. However, simulations show that either the IMA-GARCH or IMA-ARSV models offer a good alternative to approximate the dynamics of the conditionally heteroscedastic local level models.

As a general conclusion, we show that if ε_t and η_t are assumed to follow GARCH or SV processes, the conditional heteroscedasticity of the ARIMA noise a_t is weaker than the one present in the unobserved disturbances. In some cases, a_t could even be seen as homoscedastic. Therefore, the heteroscedasticity is more evident in the unobserved component model and can be overlooked when working with the reduced form ARIMA model. This result could be expected as the heteroscedasticity weakens under contemporaneous aggregation of volatility processes; see, for example, Zaffaroni (2007).

The rest of this chapter is structured as follows. In Section 2.2, we derive the kurtosis and autocorrelations of squares of the stationary transformation of the local level and smooth trend models, when the disturbances are serially uncorrelated, with symmetric distributions and finite fourth order moments. Then, we particularize these results for conditionally Normal GARCH(1,1) and ARSV(1) disturbances. In Section 2.3, we derive the statistical properties of the reduced form ARIMA noise, a_t, and show the effects of assuming GARCH or ARSV processes on this noise by means of several Monte Carlo simulations. In Section 2.4, we illustrate the main findings with an empirical application. Finally, Section 2.5 summarizes the main conclusions.

2.2 Properties of the local level and smooth trend models

In this section, we derive the excess kurtosis and the acf of squared observations for the general case in which the noises of the local level and smooth trends models are assumed to be uncorrelated processes with symmetric densities and finite fourth order moments. We then particularize these results for the case in which the disturbances are GARCH(1,1) or ARSV(1) processes.

2.2.1 The local level model

Consider the local level model given by (1.2). Note first that if the noises have symmetric distributions around zero then all odd moments of Δy_t are zero. It is also straightforward to show that its variance and acf are given by

$$Var[\Delta y_t] = \sigma_\varepsilon^2(q_\eta + 2). \tag{2.4}$$

$$\rho_\tau^{\Delta y} = \begin{cases} \dfrac{-1}{q_\eta + 2} & \tau = 1, \\ 0 & \tau \geq 2, \end{cases} \tag{2.5}$$

respectively. Furthermore, assuming finite fourth order moment of both noises, the following expression of the excess kurtosis, of Δy_t is derived

$$\bar{\kappa}_{\Delta y} = \frac{q_\eta^2 \bar{\kappa}_\eta + 2\bar{\kappa}_\varepsilon + 6(\bar{\kappa}_\varepsilon + 2)\rho_1^{\varepsilon^2}}{(q_\eta + 2)^2}, \tag{2.6}$$

where $\bar{\kappa}_\varepsilon$ and $\bar{\kappa}_\eta$ are the excess kurtosis of ε_t and η_t, respectively, and $\rho_1^{\varepsilon^2}$ is the first order autocorrelation of ε_t^2. Note that the signal-to-noise ratio, q_η, plays an important role in determining the relative influence of the excess kurtosis of each noise on $\bar{\kappa}_{\Delta y}$. In the limiting cases, when $q_\eta \to \infty$, $\Delta y_t = \eta_t$ (i.e. y_t is a pure random walk process) so that $\bar{\kappa}_{\Delta y} = \bar{\kappa}_\eta$. On the other hand, as $q_\eta \to 0$, $\Delta y_t = \Delta\varepsilon_t$ (i.e. y_t is a white noise process), and consequently Δy_t is a non-invertible MA(1) process whose excess kurtosis may be different from $\bar{\kappa}_\varepsilon$ depending on the value of $\rho_1^{\varepsilon^2}$.

Finally, the acf of $(\Delta y_t)^2$ is a function of the kurtosis and autocorrelations of the two noises, which is given by

$$\rho_\tau^{(\Delta y)^2} = \frac{q_\eta^2(\bar{\kappa}_\eta + 2)\rho_\tau^{\eta^2} + (\bar{\kappa}_\varepsilon + 2)(\rho_{\tau-1}^{\varepsilon^2} + 2\rho_\tau^{\varepsilon^2} + \rho_{\tau+1}^{\varepsilon^2})}{(\bar{\kappa}_{\Delta y} + 2)(q_\eta + 2)^2}, \ \tau \geq 1. \tag{2.7}$$

Note that for Gaussian noises, $\rho_1^{(\Delta y)^2} = (q_\eta + 2)^{-2}$, which turns out to be the squared first order autocorrelation of Δy_t; see Maravall (1983). However, when assuming that ε_t and η_t are homoscedastic but not necessarily Gaussian, it is possible to see from (2.7) that $\rho_1^{(\Delta y)^2}$ may differ from $(\rho_1^{\Delta y})^2$, depending on the values of the excess kurtosis of each noise. The numerator of (2.7) is defined as a weighted sum of two factors that depend on the lag τ. The first one, $\rho_\tau^{\eta^2}$, has a weight that is a function of q_η and $\bar{\kappa}_\eta$, while the weight of the second, $\rho_{\tau-1}^{\varepsilon^2} + 2\rho_\tau^{\varepsilon^2} + \rho_{\tau+1}^{\varepsilon^2}$, depends only on $\bar{\kappa}_\varepsilon$. As long as the acf of squares of both disturbances converges to zero, each of these factors disappears as τ increases, and therefore the acf of $(\Delta y_t)^2$ also converges to zero.

The general expressions of the excess kurtosis and acf of squares in (2.6) and (2.7) can be particularized when assuming particular specifications of the noises ε_t and η_t. Consider first that they are assumed to be GARCH(1,1) models[1] given by $\varepsilon_t = \varepsilon_t^\dagger h_t^{1/2}$ and $\eta_t = \eta_t^\dagger q_t^{1/2}$, where

[1]The general expression for the kurtosis and acf of $(\Delta y_t)^2$ can be utilized in other specifications of the noises. For instance, Broto and Ruiz (2006) derive these quantities for the particular case of a local level model with GQARCH disturbances to account for asymmetries in volatility.

$\varepsilon_t^\dagger$ and $\eta_t^\dagger$ are mutually and serially independent Normal processes with zero mean and unit variance, and

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1}, \tag{2.8a}$$

$$q_t = \gamma_0 + \gamma_1 \eta_{t-1}^2 + \gamma_2 q_{t-1}, \tag{2.8b}$$

where the parameters α_0, α_1, α_2, γ_0, γ_1 and γ_2 are assumed to satisfy the usual positivity and stationarity conditions. Then, substituting $\bar{\kappa}_\varepsilon$, $\bar{\kappa}_\eta$ and $\rho_1^{\varepsilon^2}$ in (2.6) by their corresponding expressions for the GARCH(1,1) process given by $\bar{\kappa}_\varepsilon = \frac{2\alpha_1^2}{1-3\alpha_1^2-2\alpha_1\alpha_2-\alpha_2^2}$, $\bar{\kappa}_\eta = \frac{2\gamma_1^2}{1-3\gamma_1^2-2\gamma_1\gamma_2-\gamma_2^2}$ and $\rho_1^{\varepsilon^2} = \frac{\alpha_1(1-\alpha_1\alpha_2-\alpha_2^2)}{1-2\alpha_1\alpha_2-\alpha_2^2}$, we find that

$$\bar{\kappa}_{\Delta y} = \frac{3}{(q_\eta+2)^2}\left[q_\eta^2 \frac{2\gamma_1^2}{1-3\gamma_1^2-2\gamma_1\gamma_2-\gamma_2^2} + 4\frac{\alpha_1(1+\alpha_1-\alpha_1\alpha_2-\alpha_2^2)}{1-3\alpha_1^2-2\alpha_1\alpha_2-\alpha_2^2}\right]. \tag{2.9}$$

As an illustration, Figure 2.1 plots the relationship between the kurtosis of Δy_t and the persistence of the volatility of both noises, measured by $\alpha_1 + \alpha_2$ and $\gamma_1 + \gamma_2$, for different values of the signal-to-noise ratio, when $\alpha_1 = \gamma_1 = 0.1$. The values have been selected to illustrate situations with high persistence and relatively small values of α_1 and γ_1, as commonly encountered in empirical applications. Note that the slope with respect to the persistence of η_t is steeper as q_η increases, and also that varying q_η significantly affects $\bar{\kappa}_{\Delta y}$.

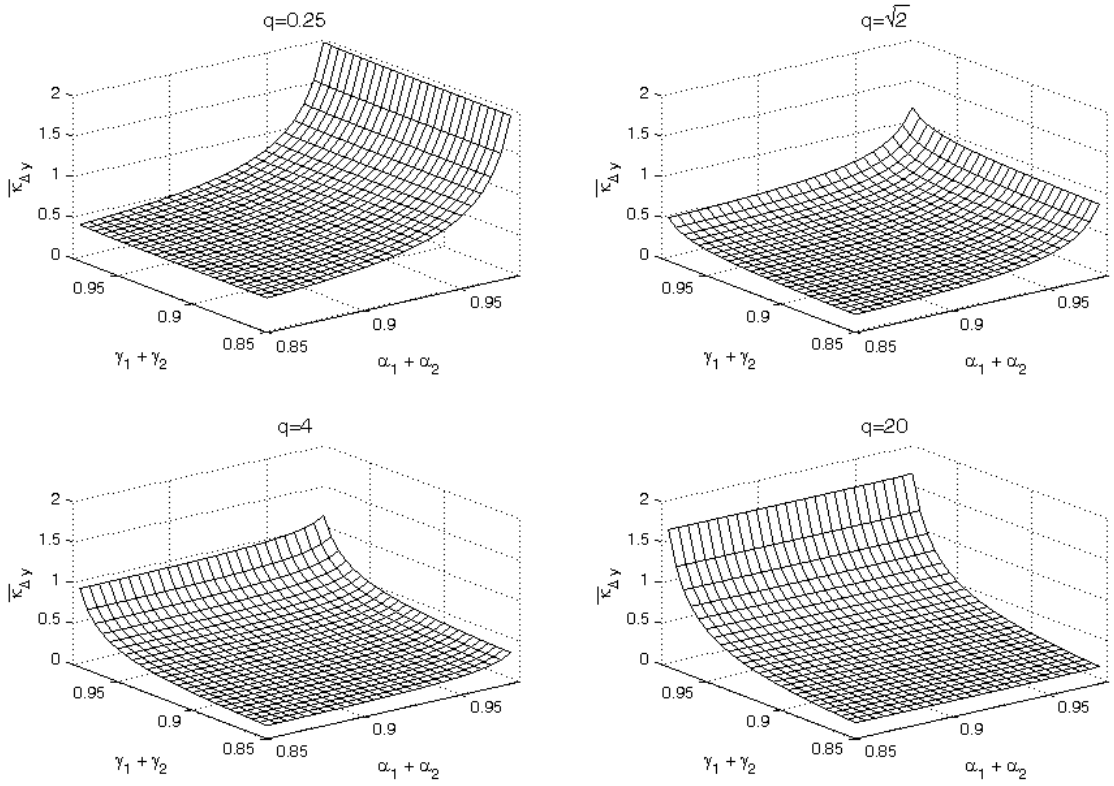

Figure 2.1: Relationship between $\bar{\kappa}_{\Delta y}$ and GARCH parameters in the local level model. The ARCH coefficients, α_1 and γ_1, are fixed to 0.10.

On the other hand, when ε_t and η_t are GARCH(1,1) processes, the acf of $(\Delta y_t)^2$ given in

(2.7) becomes

$$\rho_\tau^{(\Delta y)^2} = \begin{cases} \dfrac{q_\eta^2 \rho_1^{\eta^2}(\bar{\kappa}_\eta + 2) + (\bar{\kappa}_\varepsilon + 2)(1 + \rho_1^{\varepsilon^2}(2 + \alpha_1 + \alpha_2))}{(q_\eta + 2)^2(\bar{\kappa}_{\Delta y} + 2)}, & \tau = 1 \\ (\alpha_1 + \alpha_2)\rho_{\tau-1}^{(\Delta y)^2} + \dfrac{(\gamma_1 + \gamma_2 - \alpha_1 - \alpha_2) q_\eta^2 (\gamma_1 + \gamma_2)^{\tau-2} \rho_1^{\eta^2}(\bar{\kappa}_\eta + 2)}{(q_\eta + 2)^2(\bar{\kappa}_{\Delta y} + 2)}, & \tau \geq 2, \end{cases} \tag{2.10}$$

where $\rho_1^{\eta^2} = \frac{\gamma_1(1-\gamma_1\gamma_2-\gamma_2^2)}{1-2\gamma_1\gamma_2-\gamma_2^2}$. From (2.10) we can see that when the persistence of both noises is the same, i.e. $\gamma_1 + \gamma_2 = \alpha_1 + \alpha_2$, the acf of squares has an exponential decay, as in a GARCH(p,q) process. We can also observe an exponential decay when only one noise is heteroscedastic. However, in general, the decay of the autocorrelations in (2.10) is not exponential. Consequently, the behavior of Δy_t is not GARCH. As an illustration, Figure 2.2 plots the acf of squares for different specifications of the disturbances, together with the corresponding rates of decay from the second lag. The first row shows a model in which both disturbances follow the same GARCH process, while in the second model both noises follow GARCH processes with a different persistence. The last two rows consider models in which only one noise is heteroscedastic. Note first that the cases in the first and the last two rows illustrate the situations mentioned above where we obtain an exponential decay in the acf of $(\Delta y_t)^2$. Moreover, in the case where $\gamma_1 + \gamma_2 \neq \alpha_1 + \alpha_2$, although the rate is slightly increasing, it can be approximated by a constant. Indeed, the rate of decay of $\rho_\tau^{(\Delta y)^2}$ implicit in (2.10) converges to $max(\alpha_1 + \alpha_2; \gamma_1 + \gamma_2)$ as τ increases. In order to prove this statement, note that expression the rate of decay of $\rho_\tau^{(\Delta y)^2}$ resulting from (2.10) is given by

$$\begin{aligned} \frac{\rho_\tau^{(\Delta y)^2}}{\rho_{\tau-1}^{(\Delta y)^2}} &= \frac{A\,(\gamma_1 + \gamma_2)^{\tau-1} + B\,(\alpha_1 + \alpha_2)^{\tau-2}}{A\,(\gamma_1 + \gamma_2)^{\tau-2} + B\,(\alpha_1 + \alpha_2)^{\tau-3}} \\ &= (\alpha_1 + \alpha_2) \left[\frac{A\left(\dfrac{\gamma_1 + \gamma_2}{\alpha_1 + \alpha_2}\right)^{\tau-1} + B\,(\alpha_1 + \alpha_2)^{-1}}{A\left(\dfrac{\gamma_1 + \gamma_2}{\alpha_1 + \alpha_2}\right)^{\tau-2} + B\,(\alpha_1 + \alpha_2)^{-1}} \right], \quad \tau > 3, \end{aligned} \tag{2.11}$$

where A and B are positive constants that do not depend on τ. If we take limits to (2.11) as τ goes to infinity, we obtain two possible solutions depending on whether $\gamma_1 + \gamma_2$ is greater or smaller than $\alpha_1 + \alpha_2$. In the first case, both, the numerator and denominator of (2.11) goes to infinity as τ increases. However, by means of the L'hopital rule, we find that the limit is equal to $\gamma_1 + \gamma_2$. In the other case, when $\gamma_1 + \gamma_2 \leq \alpha_1 + \alpha_2$, the limit of (2.11) is straightforwardly obtained because many terms of the expression cancel out, yielding $\alpha_1 + \alpha_2$. Therefore, we show that the level at which the rate of decay converges is $max(\alpha_1 + \alpha_2; \gamma_1 + \gamma_2)$. This implies that in the cases where the persistence of the GARCH processes are close to each other, the rate of decay of $\rho_\tau^{(\Delta y)^2}$ will be approximately constant for almost all values of τ. Consequently, exponential structures such as the ones implied by GARCH processes can be a good approximation for the acf of squares.

Consider now that the noises, ε_t and η_t, are ARSV(1) processes given by $\varepsilon_t = \omega_\varepsilon\, \varepsilon_t^\dagger \exp\{h_t/2\}$ and $\eta_t = \omega_\eta\, \eta_t^\dagger \exp\{q_t/2\}$, with

$$h_t = \phi_h\, h_{t-1} + \nu_{h,t}\ , \tag{2.12a}$$

$$q_t = \phi_q\, q_{t-1} + \nu_{q,t}\ , \tag{2.12b}$$

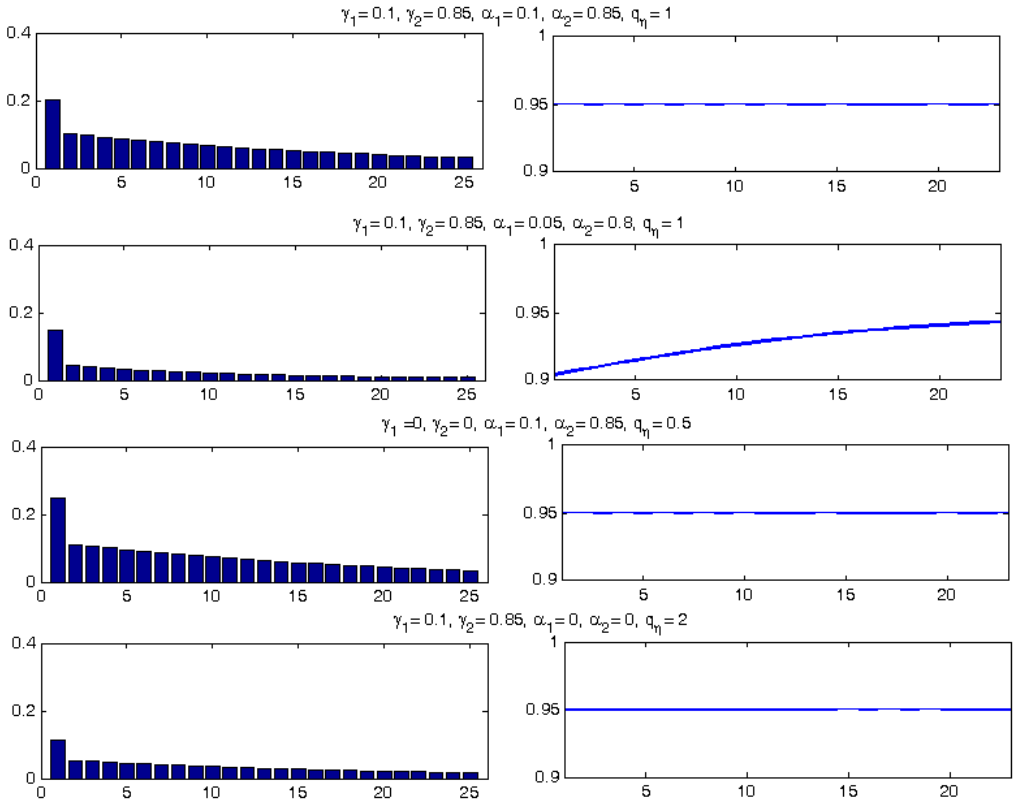

Figure 2.2: Autocorrelations of $(\Delta y_t)^2$ for several local level models with GARCH(1,1) disturbances (left column) and rates of decay defined as the ratio $\rho_\tau^{(\Delta y)^2}/\rho_{\tau-1}^{(\Delta y)^2}$ from the second lag (right column).

where $\varepsilon_t^\dagger$ and $\eta_t^\dagger$ are defined as in the GARCH case, $\omega_\varepsilon > 0$, $\omega_\eta > 0$, $|\phi_h| < 1$, $|\phi_q| < 1$, and $(\nu_{h,t}, \nu_{q,t})$ are mutually and serially uncorrelated Normal processes, independent of $(\varepsilon_t^\dagger, \eta_t^\dagger)$, with zero means and variances $\sigma_{\nu_h}^2$ and $\sigma_{\nu_q}^2$, respectively. Thus, h_t and q_t are Normally distributed with zero means and variances equal to $\sigma_h^2 = \frac{\sigma_{\nu_h}^2}{1-\phi_h^2}$ and $\sigma_q^2 = \frac{\sigma_{\nu_q}^2}{1-\phi_q^2}$, respectively. By using the properties of the log-normal distribution, we find that $\sigma_\varepsilon^2 = \omega_\varepsilon^2 \exp\{\sigma_h^2/2\}$ and $\sigma_\eta^2 = \omega_\eta^2 \exp\{\sigma_q^2/2\}$, so that we can plug them into (2.4) to obtain the marginal variance of Δy_t. Also note that, using the properties of the ARSV(1) model[2], we know that $\bar{\kappa}_\varepsilon = 3[\exp\{\sigma_h^2\} - 1]$, $\bar{\kappa}_\eta = 3[\exp\{\sigma_q^2\} - 1]$ and $\rho_1^{\varepsilon^2} = \frac{\exp\{\sigma_h^2\phi_h\}-1}{3\exp\{\sigma_h^2\}-1}$. Therefore, by plugging these moments into (2.6) we obtain the following expression of the excess kurtosis of Δy_t

$$\bar{\kappa}_{\Delta y} = \frac{3}{(q_\eta + 2)^2}\left[q_\eta^2\left(\exp\{\sigma_q^2\} - 1\right) + 2\left(\exp\{\sigma_h^2\} - 1\right) + 2\left(\exp\{\sigma_h^2\phi_h\} - 1\right)\right]. \quad (2.13)$$

Note from (2.13) that the excess kurtosis can be positive even though $\phi_h = \phi_q = 0$, provided that at least one of the variances of the processes given in (2.12) is positive. On the contrary, $\sigma_{\nu_h}^2 = \sigma_{\nu_q}^2 = 0$ is sufficient to have $\bar{\kappa}_{\Delta y} = 0$. It is also worth noting that, as long as the processes given in (2.12) are stationary, the kurtosis of Δy_t will be finite. This is not the case when the noises are GARCH, because we need further restrictions in the parameter space to ensure $\bar{\kappa}_{\Delta y} < \infty$. Figure 2.3 plots the excess kurtosis against the autoregressive parameters, ϕ_h and ϕ_q, which may be seen as persistence measures in the ARSV models. We fix $\sigma_{\nu_h}^2 = \sigma_{\nu_q}^2 = 0.05$

[2]See, for example, Taylor (1986) for a detailed study of the statistical properties of the ARSV(1) model.

and use the same values of q_η as in the GARCH case to make a fair comparison between the two types of volatility processes. As a general result, note that the behavior of $\bar{\kappa}_{\Delta y}$ is very similar to that of the GARCH case, increasing as ϕ_h and ϕ_q approach to 1, but at a rate that depends on q_η.

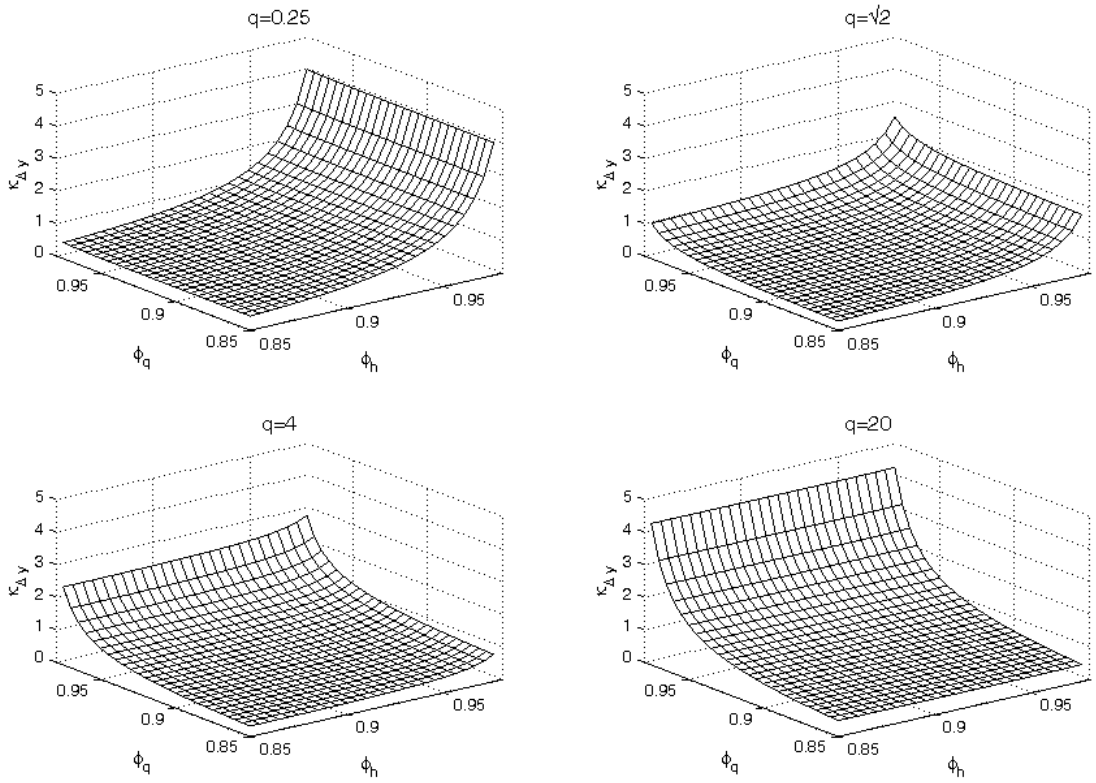

Figure 2.3: Relationship between $\bar{\kappa}_{\Delta y}$ and the persistence of both noises measured by ϕ_h and ϕ_q in the AR-SV(1) model. The variances of the log-volatility processes, $\sigma^2_{\nu_h}$ and $\sigma^2_{\nu_q}$, are fixed to 0.05.

Given that, in the ARSV(1) model, $\rho^{\varepsilon^2}_\tau = \frac{\exp\{\sigma^2_h\phi^\tau_h\}-1}{3\exp\{\sigma^2_h\}-1}$ and $\rho^{\eta^2}_\tau = \frac{\exp\{\sigma^2_q\phi^\tau_q\}-1}{3\exp\{\sigma^2_q\}-1}$, the acf of $(\Delta y_t)^2$ is given by

$$\rho^{(\Delta y)^2}_\tau = \begin{cases} \dfrac{q^2_\eta \exp\{\sigma^2_q\phi_q\} + 3\exp\{\sigma^2_h\} + 2\exp\{\sigma^2_h\phi_h\} + \exp\{\sigma^2_h\phi^2_h\} - (q^2_\eta+4)}{(q_\eta+2)^2(\bar{\kappa}_{\Delta y}+2)}, & \tau = 1 \\ \dfrac{q^2_\eta \exp\{\sigma^2_q\phi^\tau_q\} + \exp\{\sigma^2_h\phi^{\tau-1}_h\} + 2\exp\{\sigma^2_h\phi^\tau_h\} + \exp\{\sigma^2_h\phi^{\tau+1}_h\} - (q^2_\eta+4)}{(q_\eta+2)^2(\bar{\kappa}_{\Delta y}+2)}, & \tau \geq 2. \end{cases} \tag{2.14}$$

Note from (2.14) that $\sigma^2_{\nu_h} = \sigma^2_{\nu_q} = 0$ again leads to the Gaussian homoscedastic results where $\rho^{(\Delta y)^2}_\tau = \left(\rho^{\Delta y}_\tau\right)^2$. On the other hand, unlike the GARCH(1,1) process, the ARSV(1) does not produce acf of squares with an exponential decay. Figure 2.4 plots the acf of $(\Delta y_t)^2$ and the corresponding rates of decay from the second lag for four selected models. In all cases, the rates of decay are increasing and have an asymptote that is a function of the autoregressive parameters in the ARSV processes.

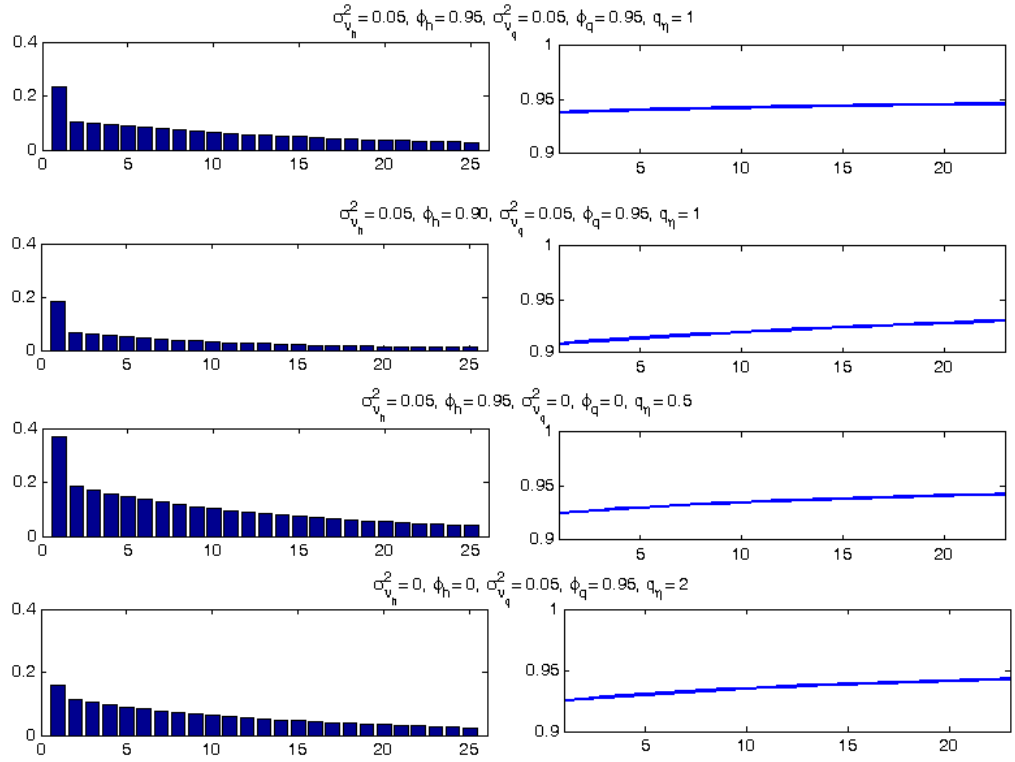

Figure 2.4: Autocorrelations of $(\Delta y_t)^2$ for several local level models with ARSV(1) disturbances (left column) and rates of decay defined as the ratio $\rho_\tau^{(\Delta y)^2}/\rho_{\tau-1}^{(\Delta y)^2}$ from the second lag (left column).

2.2.2 The smooth trend model

The variance and acf of the stationary transformation of the smooth trend model given in (1.5) are given by

$$Var[\Delta^2 y_t] = \sigma_\varepsilon^2(q_\xi + 6), \tag{2.15}$$

$$\rho_\tau^{\Delta^2 y} = \begin{cases} \dfrac{-4}{q_\xi+6} & \tau = 1, \\ \dfrac{1}{q_\xi+6} & \tau = 2, \\ 0 & \tau \geq 3, \end{cases} \tag{2.16}$$

respectively. After some tedious but straightforward algebra we find the excess kurtosis given by

$$\bar{\kappa}_{\Delta^2 y} = \frac{q_\xi^2\bar{\kappa}_\xi + 18\bar{\kappa}_\varepsilon + 6(\bar{\kappa}_\varepsilon+2)(8\rho_1^{\varepsilon^2}+\rho_2^{\varepsilon^2})}{(q_\xi+6)^2}. \tag{2.17}$$

Furthermore, the acf of $(\Delta^2 y)^2$ is given by

$$\rho_\tau^{(\Delta^2 y)^2} = \begin{cases} \dfrac{q_\xi^2(\bar{\kappa}_\xi+2)\rho_1^{\xi^2} + (\bar{\kappa}_\varepsilon+2)(8+35\rho_1^{\varepsilon^2}+8\rho_2^{\varepsilon^2}+\rho_3^{\varepsilon^2})}{(\bar{\kappa}_{\Delta^2 y}+2)(q_\xi+6)^2}, & \tau = 1, \\ \dfrac{q_\xi^2(\bar{\kappa}_\xi+2)\rho_\tau^{\xi^2} + (\bar{\kappa}_\varepsilon+2)(\rho_{\tau-2}^{\varepsilon^2}+8\rho_{\tau-1}^{\varepsilon^2}+18\rho_\tau^{\varepsilon^2}+8\rho_{\tau+1}^{\varepsilon^2}+\rho_{\tau+2}^{\varepsilon^2})}{(\bar{\kappa}_{\Delta^2 y}+2)(q_\xi+6)^2}, & \tau \geq 2. \end{cases} \tag{2.18}$$

The smooth trend model assumes that the slope of the trend evolves smoothly. Therefore, it seems sensible to assume that its noise is homoscedastic. Consequently, we only consider the possibility of the transitory noise being conditionally heteroscedastic. Moreover, given that the results above show that the conclusions for GARCH and ARSV noises are similar, in this subsection, we only consider the GARCH(1,1) process as defined in (2.8a). In this case, the excess kurtosis and acf of $(\Delta^2 y_t)^2$ are given by

$$\bar{\kappa}_{\Delta^2 y} = \frac{18\bar{\kappa}_\varepsilon + 6(\bar{\kappa}_\varepsilon + 2)(8\rho_1^{\varepsilon^2} + \rho_2^{\varepsilon^2})}{(6 + q_\xi)^2}, \tag{2.19}$$

$$\rho_\tau^{(\Delta^2 y)^2} = \begin{cases} \dfrac{(\bar{\kappa}_\varepsilon + 2)(8 + 35\rho_1^{\varepsilon^2} + 8\rho_2^{\varepsilon^2} + \rho_3^{\varepsilon^2})}{(\bar{\kappa}_{\Delta^2 y} + 2)(q_\xi + 6)^2}, & \tau = 1 \\ \dfrac{(\bar{\kappa}_\varepsilon + 2)(\rho_{\tau-2}^{\varepsilon^2} + 8\rho_{\tau-1}^{\varepsilon^2} + 18\rho_\tau^{\varepsilon^2} + 8\rho_{\tau+1}^{\varepsilon^2} + \rho_{\tau+2}^{\varepsilon^2})}{(\bar{\kappa}_{\Delta^2 y} + 2)(q_\xi + 6)^2}, & \tau = 2, 3 \\ (\alpha_1 + \alpha_2)\rho_{\tau-1}^{(\Delta^2 y)^2}, & \tau \geq 4. \end{cases} \tag{2.20}$$

Expression (2.20) implies that models like the GARCH(p,q), which produce exponential structures in the acf squares, could be suitable for the stationary transformation of the smooth trend model with a GARCH(1,1) transitory component. As an illustration, Figure 2.5 plots the acf of squares and its rate of decay for different GARCH parameters and different q_ξ. Note that the rates are all constant from the third lag, as implied in (2.20).

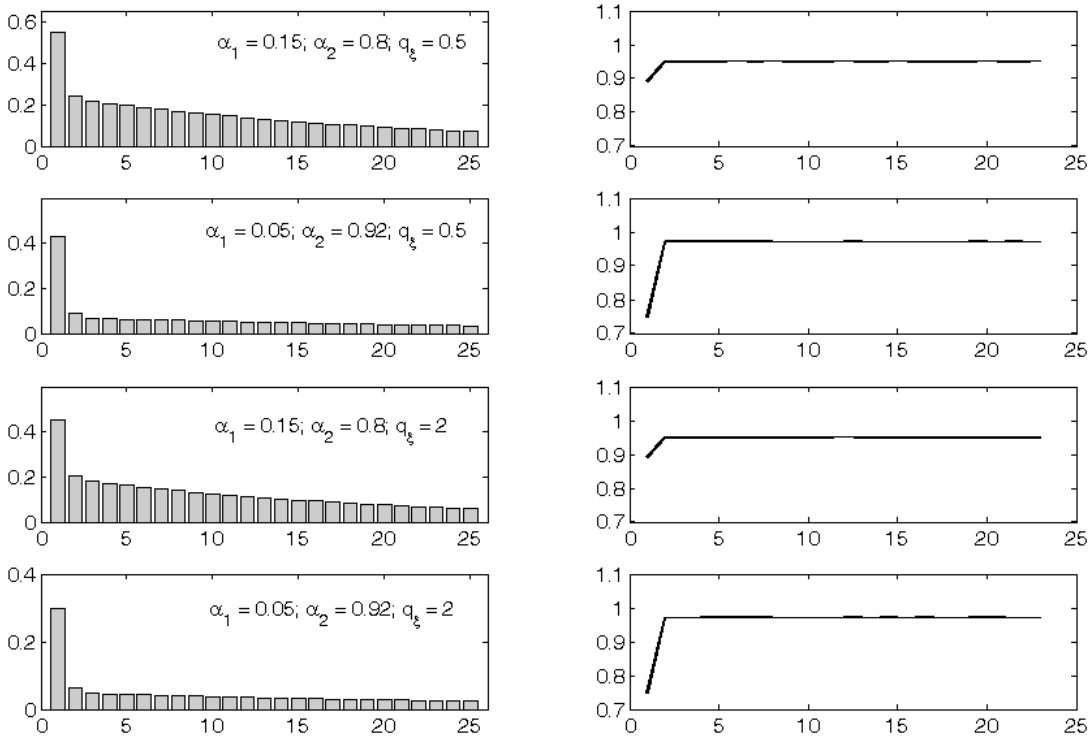

Figure 2.5: Autocorrelations of $(\Delta^2 y_t)^2$ for several smooth trend models with conditionally Normal GARCH(1,1) disturbances (left column) and rates of decay defined as the ratio $\rho_\tau^{(\Delta y)^2}/\rho_{\tau-1}^{(\Delta y)^2}$ from the third lag (right column).

2.3 Properties of the IMA noise

It is well known that ε_t and η_t being mutually and serially uncorrelated with finite positive variances is sufficient to prove that the reduced forms of the local level and smooth trend models are restricted IMA models, in the sense that both models share the same acf. Consequently, the reduced form noise, a_t, is also serially uncorrelated. Furthermore, if the disturbances of the two unobserved component models are fourth-moment stationary and have symmetric distributions, then these properties are also shared by a_t. Taking this into account, the objective of this section is twofold. First, we want to analyze the excess kurtosis and acf of squares of a_t when the disturbances of the unobserved component models are conditionally heteroscedastic. Second, by means of several Monte Carlo experiments, we study whether the GARCH and ARSV(1) models fitted to a_t are able to replicate the dynamics of the conditionally heteroscedastic unobserved component models considered here.

2.3.1 The reduced form IMA noise in the local level model

Consider the reduced form IMA(1,1) model given in (2.1). The variance and acf of Δy_t in this case are given by

$$Var[\Delta y_t] = \sigma_a^2(1+\theta^2) \tag{2.21}$$

$$\rho_\tau^{\Delta y} = \begin{cases} \dfrac{\theta}{1+\theta^2} & \tau = 1. \\ 0 & \tau \geq 2. \end{cases} \tag{2.22}$$

The excess kurtosis of Δy_t is now given by

$$\bar{\kappa}_{\Delta y} = \frac{\bar{\kappa}_a(1+\theta^4) + 6\theta^2\rho_1^{a^2}(\bar{\kappa}_a+2)}{(1+\theta^2)^2}, \tag{2.23}$$

where $\bar{\kappa}_a$ and $\rho_1^{a^2}$ are the excess kurtosis of a_t and the first order autocorrelation of a_t^2, respectively. On the other hand, it is easy to show that the acf of Δy_t^2 is given by

$$\rho_\tau^{(\Delta y)^2} = \frac{\bar{\kappa}_a+2}{(1+\theta^2)^2(\bar{\kappa}_{\Delta y}+2)}\left[(1+\theta^4)\rho_\tau^{a^2} + \theta^2(\rho_{\tau-1}^{a^2}+\rho_{\tau+1}^{a^2})\right],\ \tau \geq 1. \tag{2.24}$$

The expressions of $\bar{\kappa}_a$ and $\rho_\tau^{a^2}$ are related to those of the unobserved component noises, but in a way that is not easy to derive analytically. However, we may find approximations of these quantities by equalling the excess kurtosis of Δy_t given by (2.6) and (2.23), and the

autocorrelations of order $\tau = 1, 2, \ldots$ in (2.7) and (2.24). The following system is then obtained[3]

$$(\bar{\kappa}_a + 2)\left(1 + \theta^4 + 6\theta^2\rho_1^{a^2}\right) \equiv (1+\theta)^4(\bar{\kappa}_\eta + 2) - 8\theta(1+\theta)^2 + 2\theta^2(\bar{\kappa}_\varepsilon + 2)\left(1 + 3\rho_1^{\varepsilon^2}\right), \quad (2.25a)$$

$$(\bar{\kappa}_a + 2)\left[(1+\theta^4)\rho_\tau^{a^2} + \theta^2\left(\rho_{\tau-1}^{a^2} + \rho_{\tau+1}^{a^2}\right)\right] \equiv \theta^2(\bar{\kappa}_\varepsilon + 2)\left(\rho_{\tau-1}^{\varepsilon^2} + 2\rho_\tau^{\varepsilon^2} + \rho_{\tau+1}^{\varepsilon^2}\right) + (1+\theta)^4(\bar{\kappa}_\eta + 2)\rho_\tau^{\eta^2}, \ \tau \geq 1. \quad (2.25b)$$

From the system given in (2.25) we are able to obtain some information about the behavior of a_t. First, consider the case of homoscedastic although non-Gaussian local level disturbances, i.e. $\rho_\tau^{\varepsilon^2} = \rho_\tau^{\eta^2} = 0\ \forall\tau$ and $\bar{\kappa}_\varepsilon$ and $\bar{\kappa}_\eta$ different from zero. In this case, the autocorrelations of $(\Delta y_t)^2$ differ from the squared autocorrelations of Δy_t. Therefore, a_t is still uncorrelated but not independent; see Breidt and Davis (1992). As an illustration, Table 2.1 reports the acf of a_t^2 for several values of q_η, $\bar{\kappa}_\varepsilon$ and $\bar{\kappa}_\eta$, obtained from the resolution of the system given by (2.25). Observe that non-normality in either or both noises may generate non-zero autocorrelations of squares (specially in the first order autocorrelation). Although these autocorrelations do not follow any specific pattern and, consequently, they do not reflect the presence of GARCH or ARSV effects in the series, it is possible to obtain values of some popular conditional homoscedasticity statistics, e.g. the McLeod and Li (1983) test, that may wrongly lead to reject the null of conditionally homoscedasticity. Note also that the pattern in which only the first order autocorrelation of squares is different from zero may be confused with the effect of outliers; see Carnero et al. (2006).

q	$\bar{\kappa}_\varepsilon$	$\bar{\kappa}_\eta$	θ	$\bar{\kappa}_{\Delta y}$	$\rho_1^{(\Delta y)^2}$	$\bar{\kappa}_a$	$\rho_1^{a^2}$	$\rho_2^{a^2}$	$\rho_3^{a^2}$	$\rho_4^{a^2}$	$\rho_5^{a^2}$
0.5	0	3	-0.5	0.120	0.151	0.273	-0.030	0.008	-0.002	0.001	0.000
$\sqrt{2}$	0	3	-0.324	0.515	0.068	0.665	-0.026	0.003	0.000	0.000	0.000
0.5	3	3	-0.5	1.080	0.260	0.818	0.194	-0.048	0.012	-0.003	0.001
$\sqrt{2}$	3	3	-0.324	1.029	0.142	1.120	0.063	-0.007	0.001	0.000	0.000
0.5	3	0	-0.5	0.960	0.270	0.546	0.241	-0.060	0.015	0.004	0.001
$\sqrt{2}$	3	0	-0.324	0.515	0.171	0.456	0.109	-0.011	0.001	0.000	0.000

Table 2.1: Approximated moments of a_t resulting from local level models with either or both non-Gaussian homoscedastic noises.

Consider now that either or both noises are heteroscedastic in the form of stationary GARCH or ARSV processes as defined in the previous sections. Then, the right hand side of (2.25b)

[3] To obtain (2.25), recall that θ can be defined in terms of q_η, so that the following expressions result:

$$1 + \theta^2 = -\theta(q_\eta + 2),$$

$$1 + \theta^4 = \theta^2(q_\eta^2 + 4q_\eta + 2).$$

converges to zero as τ increases. This implies that there exists a value of τ, say τ_{max}, large enough to also make $\rho_\tau^{a^2} \approx 0$ for $\tau > \tau_{max}$. Taking this into account and solving the system backwards we can find the kurtosis and acf of squares of a_t for different specifications of the unobserved component noises. As an illustration, Figures 2.6 and 2.7 plot the acf of a_t^2 when the noises are GARCH(1,1) and ARSV(1), respectively. In general, the magnitude of the autocorrelations of a_t^2 is smaller than those of the disturbances of the local level model. This suggests that working with the reduced form of an unobserved component model may hide part of the heteroscedasticity of each component, by producing a reduced form disturbance, a_t, with less structure in its acf of squares. For instance, it might be the case that if the permanent component, μ_t, presents a significant heteroscedastic structure but the transitory component, ε_t, is homoscedastic; the stationary transformation, Δy_t, may not provide significant evidence of heteroscedasticity at all.

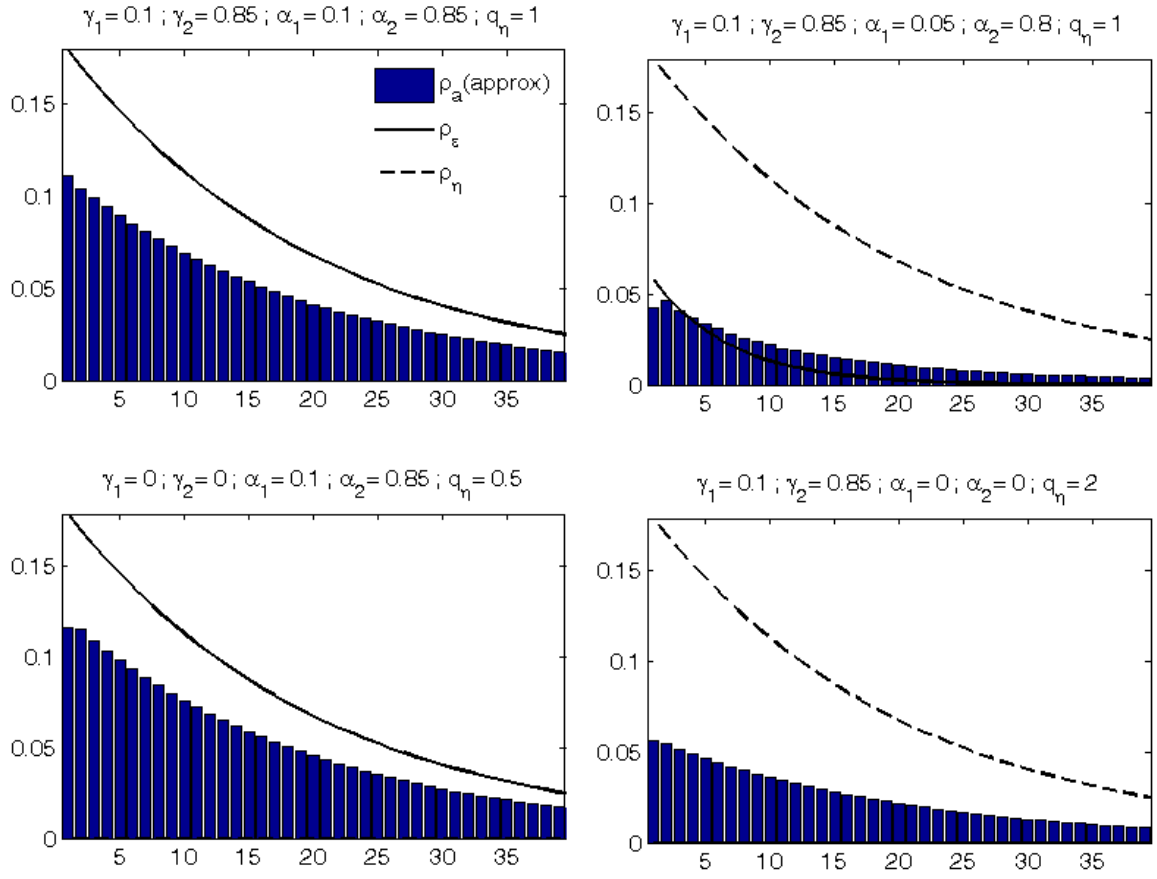

Figure 2.6: Approximated values of the acf of a_t^2 resulting from different local level models with GARCH(1,1) disturbances. The solid and dash-dotted lines draw the autocorrelations of ε_t^2 and η_t^2, respectively.

2.3.2 The reduced form IMA noise in the smooth trend model

The reduced form of the smooth trend is an IMA(2,2) as given in (1.5) with the MA parameters resulting from solving the system of equations given in (2.3). As Figure 2.8 shows, these parameters lay close to the non-invertibility frontier, specially for small values of q_ξ, which makes the IMA model very restrictive. In any case, as in the local level model, we can work out the excess

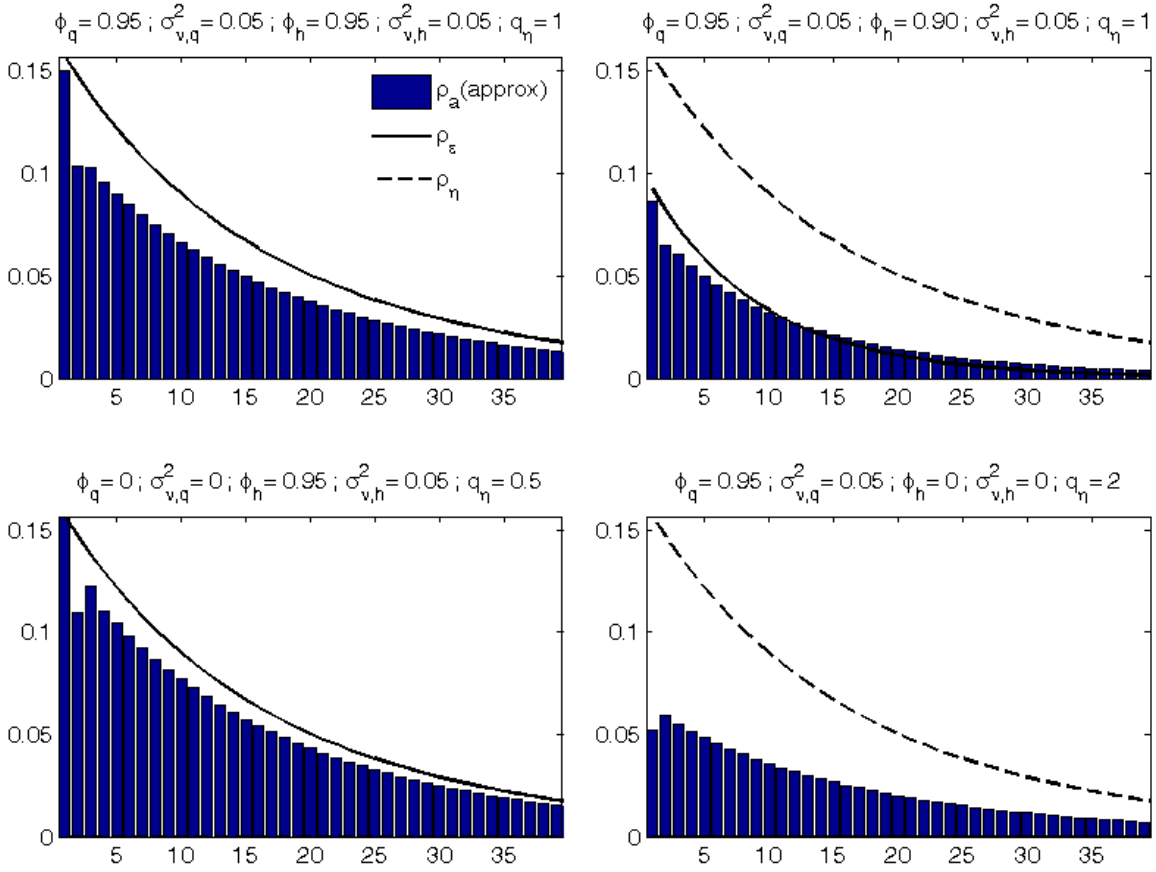

Figure 2.7: Approximated values of the acf of a_t^2 resulting from different local level models with ARSV(1) disturbances. The solid and dash-dotted lines draw the autocorrelations of ε_t^2 and η_t^2, respectively.

kurtosis and autocovariances of squares of the $\Delta^2 y_t$ as follows

$$\bar{\kappa}_{\Delta^2 y} = \frac{\bar{\kappa}_a \left(1 + \theta_1^4 + \theta_2^4\right) + 6(\bar{\kappa}_a + 2)\left[\theta_1^2 \left(1 + \theta_2^2\right) \rho_1^{a^2} + \theta_2^2 \, \rho_2^{a^2}\right]}{(1 + \theta_1^2 + \theta_2^2)^2} \tag{2.26}$$

$$\rho_\tau^{(\Delta^2 y)^2} = \begin{cases} \dfrac{\bar{\kappa}_a + 2}{(1 + \theta_1^2 + \theta_2^2)^2 (\bar{\kappa}_{\Delta^2 y} + 2)} \left[\theta_1^2 \left(1 + \theta_2^2\right) + \left(1 + \theta_1^4 + \theta_2^4 + \theta_2^2 + 4\theta_1^2 \, \theta_2\right) \rho_1^{a^2} + \right. \\ \qquad \left. \theta_1^2 \left(1 + \theta_2^2\right) \rho_2^{a^2} + \theta_2^2 \rho_3^{a^2} + \frac{4\theta_1^2 \, \theta_2}{\bar{\kappa}_a + 2}\right], \quad \tau = 1 \\ \dfrac{\bar{\kappa}_a + 2}{(1 + \theta_1^2 + \theta_2^2)^2 (\bar{\kappa}_{\Delta^2 y} + 2)} \left[\theta_2^2 \, \rho_{\tau-2}^{a^2} + \theta_1^2 \left(1 + \theta_2^2\right) \rho_{\tau-1}^{a^2} + \right. \\ \qquad \left. \left(1 + \theta_1^4 + \theta_2^4\right) \rho_\tau^{a^2} + \theta_1^2 \left(1 + \theta_2^2\right) \rho_{\tau+1}^{a^2} + \theta_2^2 \rho_{\tau+2}^{a^2}\right], \quad \tau \geq 2, \end{cases} \tag{2.27}$$

respectively. If ε_t is assumed to be GARCH(1,1), then the conditional heteroscedasticity should be also present in the resulting a_t. In this case, finding the implied acf of a_t^2 in terms of q_ξ by equating the expressions of the kurtosis in (2.17) and (2.26), and the expressions of the autocorrelations of squares in (2.18) and (2.20), is very complicate. Consequently, Figure 2.9 shows the mean estimates of the sample acf of simulated series. Looking at the plots we find the same patterns as the ones derived analytically for the local level model. The autocorrelation structure of the squared innovations is markedly weaker than that of the transitory component. As expected, the difference between these two autocorrelation functions is higher as q_ξ increases (σ_ε^2

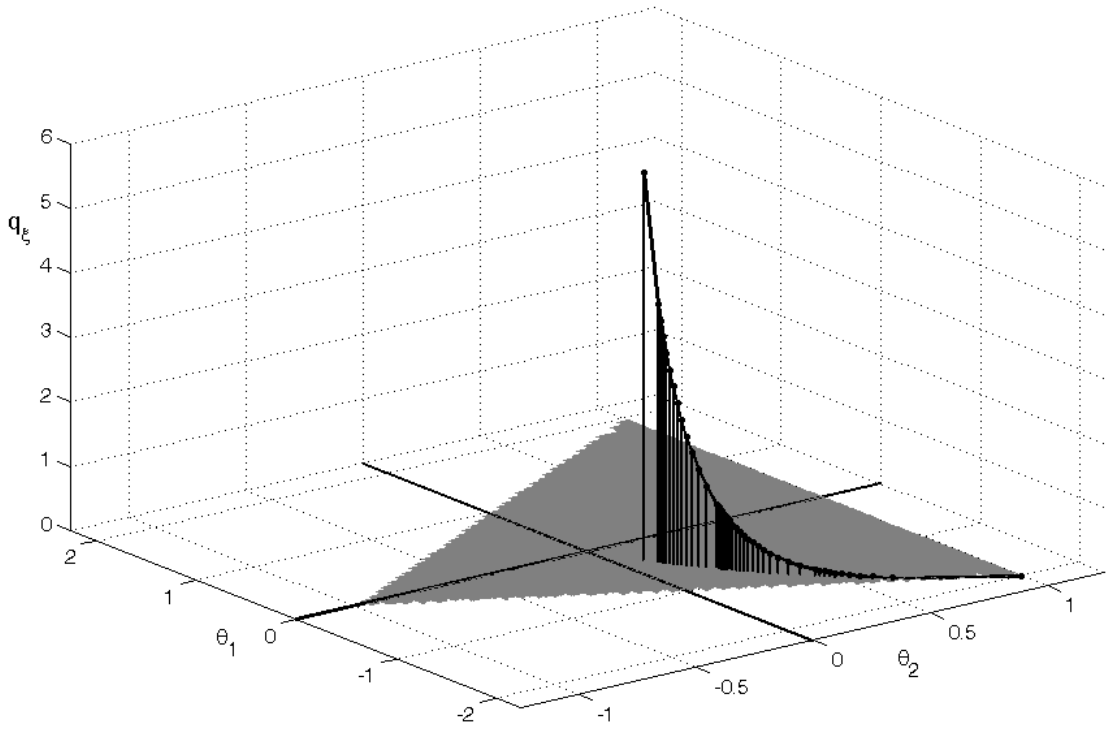

Figure 2.8: MA(2) parameters, θ_1 and θ_2, as a function of the smooth-trend parameter q_ξ (thick line). The triangular surface conforms the invertibility region of the MA(2) model.

decreases relative to σ_ξ^2). However, it seems to be invariant for different GARCH specifications. Overall, we find the same results as in the local level model. That is, the ARCH effects in the resulting ARIMA disturbance are less evident than in the unobserved components.

2.3.3 Heteroscedastic IMA models

In the previous subsection, we have seen that if the unobserved component disturbances follow either GARCH or ARSV processes, the noise of the corresponding IMA model does not share all the properties of these processes, but still shows an excess kurtosis and acf of squares different from zero. On the other hand, when analyzing real time series, it is usual to fit GARCH or ARSV processes to the residuals of an ARIMA model whenever they show evidence of conditional heteroscedasticity; see Li et al. (2002) for a review on the properties and applications of ARMA models with GARCH errors. Then, it may be relevant to study the effects of fitting IMA-GARCH and IMA-ARSV models to a series with conditionally heteroscedastic stochastic levels.

Consider first that a_t is assumed to be a conditionally Normal GARCH(1,1) model. Then $a_t = a_t^\dagger \sqrt{s_t}$, where $a_t^\dagger$ is a Gaussian white noise process and

$$s_t = \delta_0 + \delta_1 a_{t-1}^2 + \delta_2 s_{t-1}. \tag{2.28}$$

One way of approximating the parameters given in (2.28) as functions of the local level parameters is by using the relationship between κ_a and $\rho_\tau^{a^2}$ and δ_1 and δ_2 implied by the GARCH(1,1).

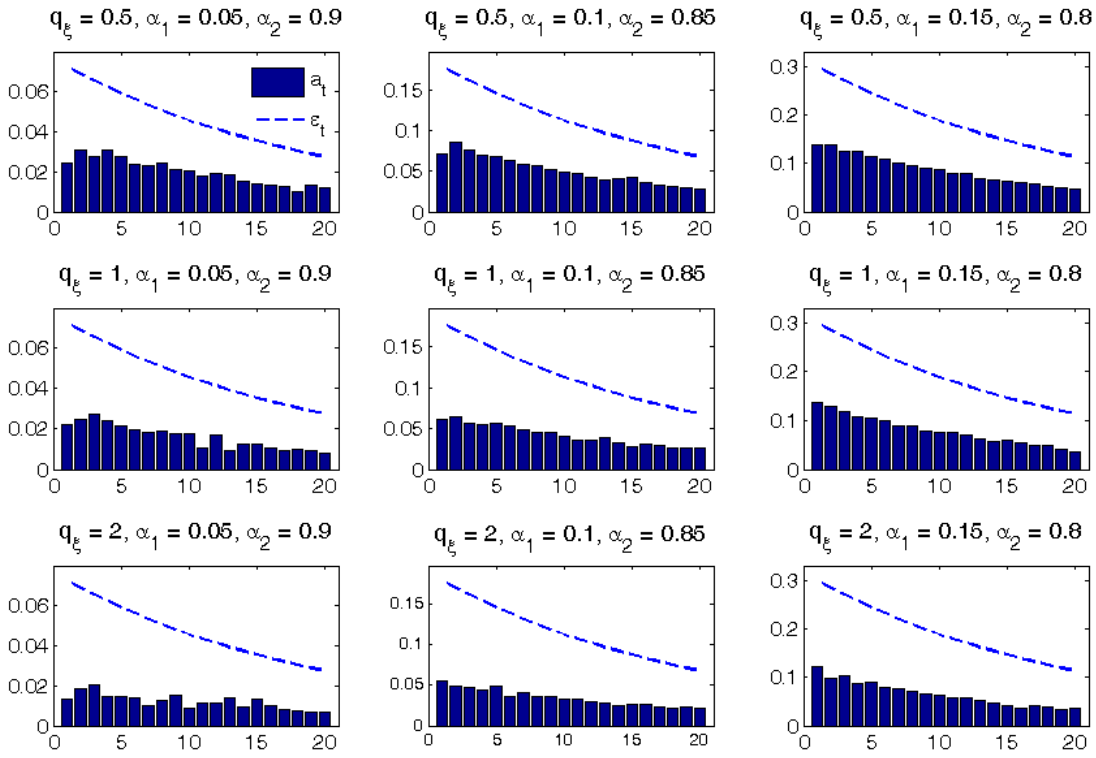

Figure 2.9: Simulated autocorrelations of a_t^2 resulting from different smooth trend models with ε_t being GARCH(1,1).

First, note that in the GARCH(1,1) model, κ_a and $\rho_\tau^{a^2}$ are given by

$$\kappa_a = \frac{3\left[1-(\delta_1+\delta_2)^2\right]}{1-3\delta_1^2-2\delta_1\delta_2-\delta_2^2}, \tag{2.29}$$

$$\rho_\tau^{a^2} = (\delta_1+\delta_2)^{\tau-1}\frac{\delta_1(1-\delta_1\delta_2-\delta_2^2)}{1-2\delta_1\delta_2-\delta_2^2},\ \tau\geq 1. \tag{2.30}$$

From (2.29) and (2.30) it is possible to derive the following expression of δ_1,

$$\delta_1 = \frac{3(\kappa_a-1)\rho_\tau^{a^2}-(\delta_1+\delta_2)^\tau(\kappa_a-3)}{2\kappa_a(\delta_1+\delta_2)^{\tau-1}}, \tag{2.31}$$

for any $\tau \geq 1$. Thus, given that we can find approximations of κ_a and $\rho_\tau^{a^2}$, by means of (2.25), then we can also find values of δ_1 and δ_2 for different values of the local level parameters. Table 2.2 shows some examples. For instance, consider the model where both noises are heteroscedastic and $q=\sqrt{2}$. In this case, (2.25) yields $\kappa_a = 4.45$, $\rho_2^{a^2} = 0.19$, $\rho_3^{a^2} = 0.18$ and $\delta_1+\delta_2 = 0.95$, so that $\delta_1 = 0.083$ (the upper-left case of Figure 2.6). This value, which measures the level of heteroscedasticity in a series, is clearly smaller than the corresponding values for ε_t and η_t, $\alpha_1 = \gamma_1 = 0.15$. If we now consider the model in which only the trend component is heteroscedastic, the values are given by $\kappa_a = 3.4$, $\rho_2^{a^2} = 0.094$, $\rho_3^{a^2} = 0.089$ and $\delta_1+\delta_2 = 0.95$, thus obtaining $\delta_1 = 0.05$. In this case, the ARCH coefficient in the reduced form disturbance is one-third the value of $\gamma_1 = 0.15$. Furthermore, if in this case, we change the value of q from 0.5 to $\sqrt{2}$ keeping the rest unchanged, we see that $\delta_1 = 0.014$, that is one-tenth of γ_1. Notice from these examples that it is possible to reject heteroscedasticity although at least one of the underlying noises is clearly heteroscedastic.

q	α_1	α_2	γ_1	γ_2	κ_ε	κ_η	θ	κ_a	$\rho_1^{a^2}$	$\rho_2^{a^2}$	$\rho_3^{a^2}$	δ_1	δ_2
0.5	0.15	0.80	0.15	0.80	5.57	5.57	-0.5	4.910	0.251	0.223	0.216	0.100	0.850
$\sqrt{2}$	0.15	0.80	0.15	0.80	5.57	5.57	-0.324	4.451	0.217	0.193	0.185	0.083	0.867
0.5	0	0	0.15	0.80	3	5.57	-0.5	3.083	0.023	0.026	0.024	0.014	0.936
$\sqrt{2}$	0	0	0.15	0.80	3	5.57	-0.324	3.396	0.092	0.094	0.089	0.049	0.901
0.5	0.15	0.80	0	0	5.57	3	-0.5	4.828	0.244	0.214	0.208	0.093	0.857
$\sqrt{2}$	0.15	0.80	0	0	5.57	3	-0.324	4.055	0.174	0.144	0.139	0.051	0.899

Table 2.2: Approximated values of the GARCH parameters, δ_1 and δ_2, of the reduced form IMA noise, a_t, corresponding to different local level models with GARCH noises.

In order to analyze the effects of fitting IMA models with GARCH noises to conditionally heteroscedastic local level and smooth trend models, we carry out several Monte Carlo simulations. For each specification of the conditionally heteroscedastic unobserved component models, we generate 1000 series and fit the IMA-GARCH model to analyze the pattern of values of δ_1 and δ_2 that we can expect to obtain in practice. The series are generated with three different sample sizes $T = 200,\ 1000$ and 5000. The parameters of the IMA-GARCH model are estimated by QML in two steps, estimating first the MA parameter, θ, and then fitting the GARCH model to the residuals. Furthermore, we also test for homoscedasticity in the residuals of the first step using the test proposed by Rodriguez and Ruiz (2005) and given by

$$Q_1(10) \quad = \quad T\sum_{k=1}^{9}\left[\tilde{r}(k)+\tilde{r}(k+1)\right]^2,$$

where $\tilde{r}(k) = \sqrt{(T+2)/(T-k)}r(k)$ is the standardized sample autocorrelation of order k. This test is more powerful than the popular McLeod-Li test. Table 2.3 reports the Monte Carlo means and standard deviations of the QML estimates, together with the percentage of rejections of the homoscedasticity in the residuals, when the series are generated by a local level model with both noises being GARCH(1,1) with strong ARCH effects. Additionally, Table 2.4 reports the same results when the series are generated by a local level model with only the noise of the permanent component being GARCH(1,1). Note first that the estimates of the MA parameter are unbiased even in moderate sample sizes. On the other hand, note that when $T = 200$, we do not reject the null of homoscedasticity in the residuals in a large proportion of cases, namely 47.6% for the case in which both noises are GARCH and 65% for the case in which only η_t is GARCH. This result is also reflected in the fact that the estimates of the ARCH effect in the reduced form noise are not significantly different from zero when $T = 200$. Increasing the sample size leads to significant ARCH effects although the are clearly smaller than those of the local level disturbances. This result supports the conclusions found in Table 2.2 with the approximated values of δ_1 as functions of the local level parameters. It is also important to observe that the average estimate of the ARCH parameter is the same regardless of the sample size. However, the Monte Carlo average of the GARCH parameter increases with the sample size. Thus, in large samples, the ARCH parameter is underestimated with

respect to the ARCH parameters in the local level disturbances, while the GARCH parameter is overestimated. Note that the average persistence of the reduced form GARCH model is the same as the persistence observed in the local level disturbances. Tables 2.3 and 2.4 also report the excess kurtosis and the autocorrelations of squares of Δy_t, together with the corresponding average plug-in moments obtained when the estimated parameters are substituted in expressions (2.23) and (2.24). Comparing the moments of Δy_t with the corresponding plug-in moments, we observe that although the latter increase with the sample size, they are clearly smaller than the population moments even when $T = 5000$.

Tables 2.5 and 2.6 report the same quantities as Tables 2.3 and 2.4, but considering that the noises follow ARSV(1) processes. In general, the conclusions are similar. The MA parameter is estimated unbiasedly and, in moderate samples, there is a large percentage of series in which the homoscedasticity is rejected. In this case, the ARCH estimate even decreases with the sample size while the GARCH estimate increases. In large samples, the average persistence is close to this implied by the ARSV models. Finally, the plug-in kurtosis is much smaller than this in the population while the plug-in autocorrelations of squares only slightly underestimate their population counterparts.

In the last two exercises, we have also analyzed the properties of the reduced form IMA model when its disturbance is assumed to be ARSV(1) given by $a_t = a_t^\dagger \, \omega_a \exp\{s_t/2\}$, where $a_t^\dagger \sim NID(0,1)$ and

$$s_t = \phi_a s_{t-1} + \nu_{a,t} \tag{2.32}$$

with $\nu_{a,t} \sim NID(0, \sigma^2_{\nu,a})$ and independent of $a_t^\dagger$. Tables 2.3 and 2.5 report the average means and standard deviations of the QML estimates of ϕ_a and $\sigma^2_{\nu,a}$ as proposed by Harvey et al. (1994). The average estimates of ϕ_a increase with the sample size, while those of $\sigma^2_{\nu,a}$ are approximately constant. Note that this result is in concordance with the result obtained when the GARCH model is fitted to the MA residuals, in which the persistence converges to this in the local level noises while the ARCH parameter is underestimated and roughly constant for all the sample sizes. The estimates of $\sigma^2_{\nu,a}$ are not significant for $T = 1000$. Finally, note that the plug-in kurtosis is closer to the population kurtosis than this implied by the GARCH model, while the plug-in autocorrelations of squares of both models are similar.

Finally, we have also carried out the exercise for the smooth trend model, when ε_t is GARCH. Table 2.7 shows the results. In general, the conclusions are roughly the same as in the local level model. Note that again, the null of homoscedasticity is not rejected in a larger proportion of times, specially when $T = 200$. With respect to the plug-in moments for $T = 5000$, note that both, the IMA-GARCH and IMA-ARSV models provide an implied excess kurtosis and first order autocorrelation of squares substantially smaller than the population counterparts, implying that using these two models may reduce the nonlinearity coming from the conditional heteroscedastic transitory component.

It is possible to draw some general conclusions from the Monte Carlo experiments carried out above. First, note that the conclusions are roughly the same across all conditionally heteroscedastic unobserved component models considered in the simulations. For example, the MA

parameter is estimated accurately, even in small samples, regardless of the model used. This result is expected given that the second marginal moments of the unobserved component models do not change when imposing stationary GARCH or ARSV processes and therefore, the relationship between θ and these moments is preserved. Additionally, in relatively small samples, a weak structure in the autocorrelations of a_t^2 may lead to under-reject the null of homoscedasticity, even tough the test-statistic is more powerful than the usual McLeod-Li statistic. Moreover, even in large samples, the parameters determining the level of these autocorrelations in the GARCH and ARSV models, δ_1 and $\sigma^2_{\nu,a}$, are typically underestimated when compared with α_1 and γ_1 or $\sigma^2_{\nu,h}$ and $\sigma^2_{\nu,q}$, respectively. Finally, by looking at the excess kurtosis and autocorrelations of squares, it seems that the moments implied by the IMA-ARSV model adjusts better the former and the moments implied by the IMA-GARCH the latter, for $T = 5000$.

LL		Estimated IMA(1,1) on y_t		
Parameters		$T = 200$	$T = 1000$	$T = 5000$
$\theta = -0.382$	$\hat{\theta} =$	-0.389 (0.12)	-0.386(0.06)	-0.383(0.03)
$q = 1$	$Q_1(10) =$	52.4%	99.6%	100.0%

GARCH(1,1)		Estimated GARCH(1,1) on a_t			Estimated ARSV(1) on a_t			
Parameters		$T = 200$	$T = 1000$	$T = 5000$		$T = 200$	$T = 1000$	$T = 5000$
$\alpha_1 = 0.15$; $\gamma_1 = 0.15$	$\hat{\delta}_1 =$	0.094(0.06)	0.094(0.02)	0.094(0.01)	$\hat{\sigma}^2_{\nu,a} =$	0.362(0.88)	0.071(0.17)	0.039(0.01)
$\alpha_1 + \alpha_2 = 0.95$;	$\hat{\delta}_1 + \hat{\delta}_2 =$	0.841(0.21)	0.937(0.04)	0.945(0.01)	$\hat{\phi}_a =$	0.743(0.25)	0.913(0.09)	0.942(0.02)
$\gamma_1 + \gamma_2 = 0.95$								

LL-GARCH(1,1)		Sample and plug-in moments when $T = 200$			Sample and plug-in moments when $T = 5000$		
Moments	Population	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)
$\bar{\kappa}_{(\Delta y)} =$	1.770	0.629(1.08)	0.586(0.79)	2.296(3.26)	1.498(1.57)	0.774(0.18)	1.194(0.23)
$\rho_1^{(\Delta y)^2} =$	0.333	0.177(0.12)	0.147(0.17)	0.071(0.06)	0.264(0.06)	0.145(0.02)	0.109(0.02)

Table 2.3: Monte Carlo averages and standard deviations (in parenthesis) of the QML estimates of the IMA-GARCH and IMA-ARSV parameters and the plug-in excess kurtosis and first order autocorrelation of squares. The series are generated by a local level model with GARCH processes in both noises. $Q(10)$ reports the percentage of series in which the null of homoscedasticity is rejected at 5% when using the statistics proposed by Rodriguez and Ruiz (2005) at lag 10, with the correction proposed by Rodriguez and Ruiz (2005). The plug-in excess kurtosis ($\bar{\kappa}_{\Delta y}$) and first order autocorrelation of squares ($\rho_1^{(\Delta y)^2}$) are constructed using the corresponding expressions for both models and substituting the parameters for their QML estimates.

LL		Estimated IMA(1,1) on y_t		
Parameters		$T = 200$	$T = 1000$	$T = 5000$
$\theta = -0.268$	$\hat{\theta} =$	-0.292(0.09)	-0.272(0.05)	-0.269(0.02)
$q = 2$	$Q_1(10) =$	35.2%	92.8%	100.0%

GARCH(1,1)		Estimated GARCH(1,1) on a_t				Estimated ARSV(1) on a_t		
Parameters		$T = 200$	$T = 1000$	$T = 5000$		$T = 200$	$T = 1000$	$T = 5000$
$\alpha_1 = 0$; $\gamma_1 = 0.15$	$\hat{\delta}_1 =$	0.069(0.06)	0.066(0.03)	0.065(0.01)	$\hat{\sigma}^2_{\nu,a} =$	0.414(0.98)	0.108(0.29)	0.026(0.03)
$\alpha_1 + \alpha_2 = 0$;	$\hat{\delta}_1 + \hat{\delta}_2 =$	0.776(0.26)	0.901(0.11)	0.932(0.02)	$\hat{\phi}_a =$	0.687(0.25)	0.850(0.18)	0.934(0.05)
$\gamma_1 + \gamma_2 = 0.95$								

LL-GARCH(1,1)		Sample and plug-in moments when $T = 200$			Sample and plug-in moments when $T = 5000$		
Moments	Population	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)
$\bar{\kappa}_{(\Delta y)} =$	0.643	0.172(0.61)	0.258(0.29)	2.167(4.21)	0.547(0.80)	0.267(0.07)	0.641(0.18)
$\rho_1^{(\Delta y)^2} =$	0.177	0.110(0.10)	0.173(0.13)	0.130(0.06)	0.157(0.05)	0.146(0.01)	0.133(0.02)

Table 2.4: Monte Carlo averages and standard deviations (in parenthesis) of the QML estimates of the IMA-GARCH and IMA-ARSV parameters and the plug-in excess kurtosis and first order autocorrelation of squares. The series are generated by a local level model with a GARCH process only in the noise of the permanent component, η_t. See Table 2.3 for further detail.

LL		Estimated IMA(1,1) on y_t		
Parameters		$T = 200$	$T = 1000$	$T = 5000$
$\theta = -0.382$	$\hat{\theta} =$	-0.402(0.13)	-0.386(0.07)	-0.382(0.03)
$q = 1$	$Q_1(10) =$	88.3%	100%	100.0%

ARSV(1)		Estimated GARCH(1,1) on a_t			Estimated ARSV(1) on a_t			
Parameters		$T = 200$	$T = 1000$	$T = 5000$		$T = 200$	$T = 1000$	$T = 5000$
$\sigma^2_{\nu,h} = 0.1$; $\sigma^2_{\nu,q} = 0.1$	$\hat{\delta}_1 =$	0.155(0.09)	0.146(0.04)	0.143(0.02)	$\hat{\sigma}^2_{\nu,a} =$	0.400(0.86)	0.095(0.08)	0.075(0.02)
$\phi_h = 0.95$; $\phi_q = 0.95$	$\hat{\delta}_1 + \hat{\delta}_2 =$	0.865(0.17)	0.947(0.03)	0.955(0.01)	$\hat{\phi}_a =$	0.790(0.22)	0.922(0.05)	0.937(0.02)

LL-ARSV(1)		Sample and plug-in moments when $T = 200$			Sample and plug-in moments when $T = 5000$		
Moments	Population	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)
$\bar{\kappa}_{(\Delta y)} =$	3.516	1.572(1.58)	0.831(2.51)	3.709(4.10)	2.812(1.33)	3.530(1.83)	2.473(0.30)
$\rho_1^{(\Delta y)^2} =$	0.410	0.235(0.13)	0.260(0.24)	0.105(0.07)	0.304(0.06)	0.272(0.05)	0.158(0.01)

Table 2.5: Monte Carlo averages and standard deviations (in parenthesis) of the QML estimates of the IMA-GARCH and IMA-ARSV parameters when the series are generated by a local level model with ARSV(1) processes in both noises. See Table 2.3 for further explanation.

LL		Estimated IMA(1,1) on y_t		
Parameters		$T = 200$	$T = 1000$	$T = 5000$
$\theta = -0.268$	$\hat{\theta} =$	-0.297(0.11)	-0.278(0.05)	-0.269(0.02)
$q = 1$	$Q_1(10) =$	45.9%	98.7%	100%

ARSV(1)		Estimated GARCH(1,1) on a_t			Estimated ARSV(1) on a_t			
Parameters		$T = 200$	$T = 1000$	$T = 5000$		$T = 200$	$T = 1000$	$T = 5000$
$\sigma^2_{\nu,h} = 0$; $\sigma^2_{\nu,q} = 0.1$	$\hat{\delta}_1 =$	0.095(0.08)	0.085(0.03)	0.083(0.01)	$\hat{\sigma}^2_{\nu,a} =$	0.463(1.08)	0.122(0.31)	0.035(0.02)
$\phi_h = 0$; $\phi_q = 0.95$	$\hat{\delta}_1 + \hat{\delta}_2 =$	0.796(0.24)	0.918(0.07)	0.935(0.02)	$\hat{\phi}_a =$	0.714(0.26)	0.870(0.17)	0.934(0.03)

LL-ARSV(1)		Sample and plug-in moments when $T = 200$			Sample and plug-in moments when $T = 5000$		
Moments	Population	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)
$\bar{\kappa}_{(\Delta y)} =$	1.342	0.693(1.08)	0.465(0.56)	2.870(5.76)	1.287(0.78)	0.460(0.10)	0.878(0.19)
$\rho_1^{(\Delta y)^2} =$	0.161	0.119(0.09)	0.216(0.15)	0.147(0.06)	0.154(0.04)	0.179(0.02)	0.151(0.02)

Table 2.6: Monte Carlo averages and standard deviations (in parenthesis) of the QML estimates of the IMA-GARCH and IMA-ARSV parameters when the series are generated by a local level model with an ARSV(1) process only in the noise of the permanent component, η_t. See Table 2.3 for further explanation.

ST		Estimated IMA(2,2) on y_t		
Parameters		$T = 200$	$T = 1000$	$T = 5000$
$\theta_1 = -0.911$	$\hat{\theta}_1 =$	-0.889(0.11)	-0.904(0.05)	-0.908(0.02)
$\theta_2 = 0.295$	$\hat{\theta}_2 =$	0.277(0.09)	0.291(0.04)	0.296(0.02)
$q_\xi = 0.5$	$Q_1(10) =$	38.9%	97.7%	100.0%

GARCH(1,1)		Estimated GARCH(1,1) on a_t				Estimated ARSV(1) on a_t		
Parameters		$T = 200$	$T = 1000$	$T = 5000$		$T = 200$	$T = 1000$	$T = 5000$
$\alpha_1 = 0.15$	$\hat{\delta}_1 =$	0.084 (0.07)	0.076 (0.02)	0.077 (0.01)	$\hat{\sigma}^2_{\nu,a} =$	0.384(0.93)	0.098(0.28)	0.029(0.02)
$\alpha_1 + \alpha_2 = 0.95$;	$\hat{\delta}_1 + \hat{\delta}_2 =$	0.823(0.21)	0.926(0.05)		$\hat{\phi}_a =$	0.707(0.24)	0.871(0.16)	0.936(0.03)

ST-GARCH(1,1)		Sample and plug-in moments when $T = 200$			Sample and plug-in moments when $T = 5000$		
Moments	Population	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)	Sample	IMA-GARCH(1,1)	IMA-ARSV(1)
$\bar{\kappa}_{(\Delta^2 y)} =$	2.839	0.787(1.03)	0.641(0.58)	1.566(2.52)	2.209(1.74)	0.600(0.13)	0.762(0.15)
$\rho_1^{(\Delta^2 y)^2} =$	0.549	0.426(0.12)	0.408(0.10)	0.423(0.07)	0.517(0.05)	0.407(0.01)	0.434(0.02)

Table 2.7: Monte Carlo averages and standard deviations (in parenthesis) of the QML estimates of the IMA-GARCH and IMA-ARSV parameters when the series are generated by a smooth trend model with ε_t being GARCH. See Table 2.3 for further explanation.

2.4 An empirical illustration

In this section we illustrate some of the main results found above with a real time series. In particular, we analyze the daily Pound/Euro exchange rate defined as the daily closing price of the Pound (£) per unit of Euro (€). The series has been downloaded from the EcoWin database and the sample spans 6 years of data, from 01/03/2000 to 03/29/2006, with $T = 1626$ observations. Figure 2.10 plots the series, its first differences and the sample autocorrelations of the level and squared observations. From the graphs we see that the exchange rates follow a non stationary pattern while returns are stationary with periods of clustered volatility. Additionally, the correlogram of the first difference suggests that it can be well represented by an MA(1) model with $\theta < 0$ and, consequently, the dynamic dependence of the series of exchange rates, y_t, can be explained by the local level model. Furthermore, the significant autocorrelations of squares may indicate the presence of conditional heteroscedasticity.

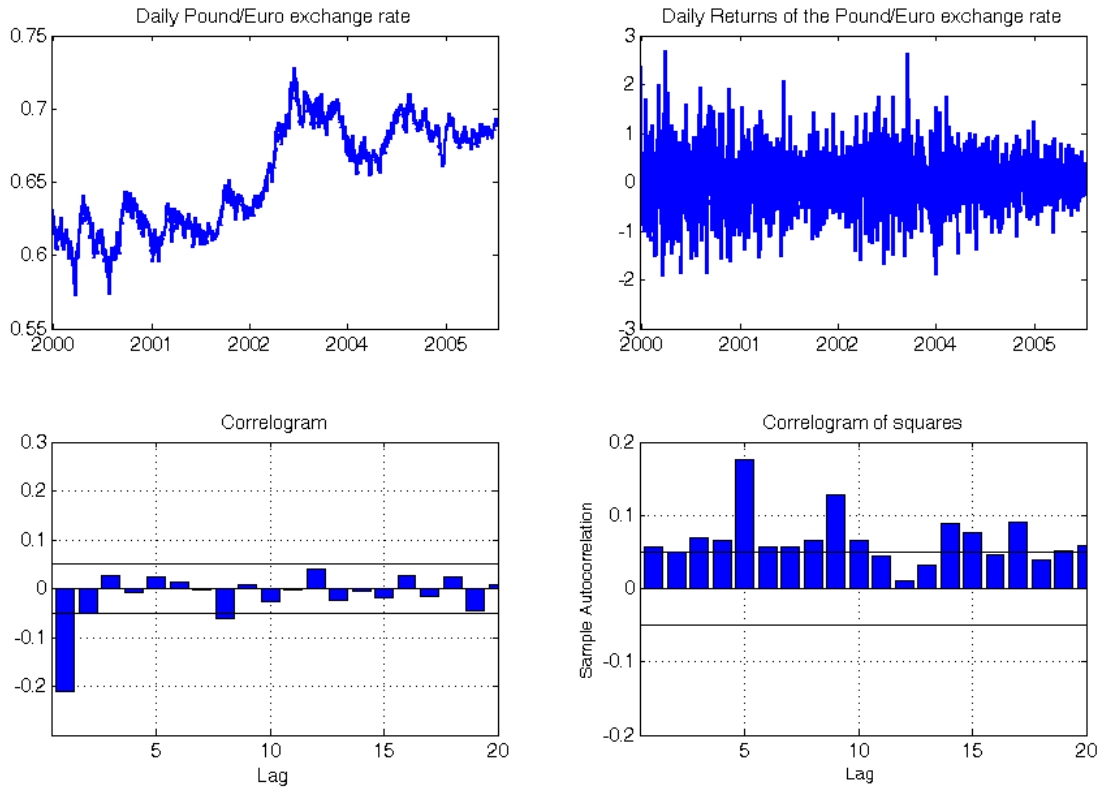

Figure 2.10: The daily £/€ exchange rate, from 01/03/2000 to 03/29/2006.

Following the procedure of Broto and Ruiz (2009), we first fit the homoscedastic unobserved component model to the series to obtain information about which component is conditionally heteroscedastic by analyzing the auxiliary residuals[4]. We also fit the IMA(1,1) model to the

[4]Prior to estimation, an intervention analysis of the series using auxiliary residuals, proposed by Harvey and Koopman (1992), was carried out with the program STAMP 6.20 of Koopman et al. (2000). The program found two outliers in the transitory component (ε_t) and three in the noise of the permanent component (η_t), representing just a 0.3% of the observations.

corresponding first differences[5]. Table 2.8 reports the estimates of the two homoscedastic models. We can see that the permanent component has more weight in the series as $\hat{q}_\eta > 1$. Furthermore, the estimate of θ (-0.247) is significant and almost identical to the one implied by $\hat{q}_\eta$ (-0.248), as expected given that the two models are still equivalent in terms of their conditional mean, regardless of the possible presence of conditional heteroscedasticity in the noises. Table 2.8 also reports the sample mean, skewness (SK), excess kurtosis ($\bar{\kappa}$), and autocorrelations of the standardized one-step-ahead residuals, $\hat{v}_t$ in the local level model, and $\hat{a}_t$ in the IMA model. They are clearly uncorrelated suggesting that the two models seem to be appropriate to fit the conditional mean. However, when we look at the sample autocorrelations of the squares, it is clear the presence of a certain structure in the variance of the series that is not captured by the homoscedastic models. This fact, in conjunction with a significant excess kurtosis, leads to propose a GARCH process to account for this structure in the variance.

Table 2.8 also reports the sample moments and autocorrelations of the auxiliary residuals and their squares in the LL model. These values are useful tools to identify which of the components present evidence of conditional heteroscedasticity. Since by construction both auxiliary residuals are serially correlated, we have to find the sample autocorrelations of squares adjusted by the square of the sample acf to account for a significant nonlinear structure in the form of conditional heteroscedasticity. Figure 2.11 shows both adjusted correlograms. From them we can conclude that both components seem to have conditional heteroscedasticity. Then, in the selection of the local level model that best captures the conditional heteroscedasticity, we should include a GARCH specification in both noises.

On the other hand, for the reduced form model selection, the sample autocorrelations given in Table 2.8 suggest that the IMA(1,1)-GARCH(1,1) could also be an adequate reduced form model to fit the series. Therefore, we fit the two models to the daily £/€ exchange rate. Table 2.9 reports the estimation results. The estimates of the LL-GARCH model imply that both noises are conditionally heteroscedastic. Furthermore, as expected, the transitory component seems to be more heteroscedastic than the permanent one as $\hat{\alpha}_1 > \hat{\gamma}_1$, although both share almost the same (high) persistence. On the other hand, compared to the homoscedastic specification, the introduction of a GARCH process in each noise increases the Log-Likelihood from -1439 to -1372. In addition to this, the residuals standardized by their estimated conditional variances not only are uncorrelated but also present almost no evidence of conditional heteroscedasticity (all the Q-statistics are insignificant at 1% and only one significant at 5%). We have to point out, however, that a small structure in the conditional variances of both noises still remains after fitting the GARCH process.

Fitting the conditionally heteroscedastic local level model allows us to obtain not only estimates of the parameters, but also of the unobserved components with their corresponding volatility. Then, it is possible to decompose the volatility of the series into the sum of the volatility of the transitory component, ($\hat{h}_t$) and the permanent component ($\hat{q}_t$). Figure 2.12 plots these two volatilities over time, as well as the volatility of the reduced form disturbance,

[5]According to both, the Akaike and Schwarz information criteria, the best fit for the Pound/Euro exchange rate is obtained with the IMA(1,1) model.

Local Level model				IMA(1,1) model	
	$\hat{\sigma}_\varepsilon^2 =$	0.084**		$\hat{\sigma}_a^2 =$	0.344**
	$\hat{\sigma}_\eta^2 =$	0.196**		$\hat{\theta} =$	-0.247**
	$\hat{q} =$	2.338			
	$\hat{\upsilon}_t$	$\hat{\varepsilon}_t$	$\hat{\eta}_t$		$\hat{a}_t$
Mean	0.016	0.000	0.019	Mean	0.016
SK	0.205	-0.002	0.240	SK	0.201
κ	3.981	3.486	4.056	κ	3.962
ρ_1	0.012	-0.353**	0.244**	ρ_1	0.014
ρ_2	-0.044	-0.148**	0.022	ρ_2	-0.043
ρ_3	0.017	0.007	0.024	ρ_3	-0.017
ρ_4	0.002	-0.017	0.015	ρ_4	-0.002
ρ_5	0.028	0.023	0.032	ρ_5	0.028
ρ_{10}	-0.031	-0.021	-0.036	ρ_{10}	-0.031
$Q(10)$	15.434	247.2**	116.28**	$Q(10)$	15.489
	$\hat{\upsilon}_t^2$	$\hat{\varepsilon}_t^2$	$\hat{\eta}_t^2$		$\hat{a}_t^2$
ρ_1	0.038	0.189**	0.092**	ρ_1	0.038
ρ_2	0.029	0.078**	0.016	ρ_2	0.029
ρ_3	0.054*	-0.094**	0.035	ρ_3	-0.054*
ρ_4	0.052*	-0.080**	0.056*	ρ_4	-0.052*
ρ_5	0.170**	0.151**	0.154**	ρ_5	0.171**
ρ_{10}	0.065**	0.087**	0.079**	ρ_{10}	0.065**
$Q_1(10)$	365.44**	352.30**	600.90**	$Q_1(10)$	369.99**
$LogL$	-1439			$LogL$	-1437

Table 2.8: Estimates and sample moments of the residuals of the homoscedastic LL and IMA(1,1) models fitted to the £/€ exchange rate. $Q(10)$ and $Q_1(10)$ are the Ljung-Box and the Rodriguez and Ruiz (2005) test statistics to test for the 10 first autocorrelations being jointly zero. *(**) Significant at 5% (1%) level.

a_t. All these volatilities show a common pattern. First, at the beginning of the Euro as a common currency, the uncertainty about its behavior leads to a highly volatile period. Then it begins to decrease until the end of 2003, where there is a new increase of the uncertainty surrounding the exchange rate. Finally, the last two years show a smooth decreasing pattern. By construction, the information given by the reduced form disturbance, a_t, cannot provide any extra information about the sources of these highly volatile and quiet periods. However, from the volatility of both structural noises, we observe that the contribution of each component to the total volatility has been different throughout the sample. While the first highly volatile period is almost totally driven by the permanent component, the source of volatility in the second period is shared by the two components. However, we can see that the decreasing behavior of the last two years is accompanied by a gradual reduction in the contribution of the transitory

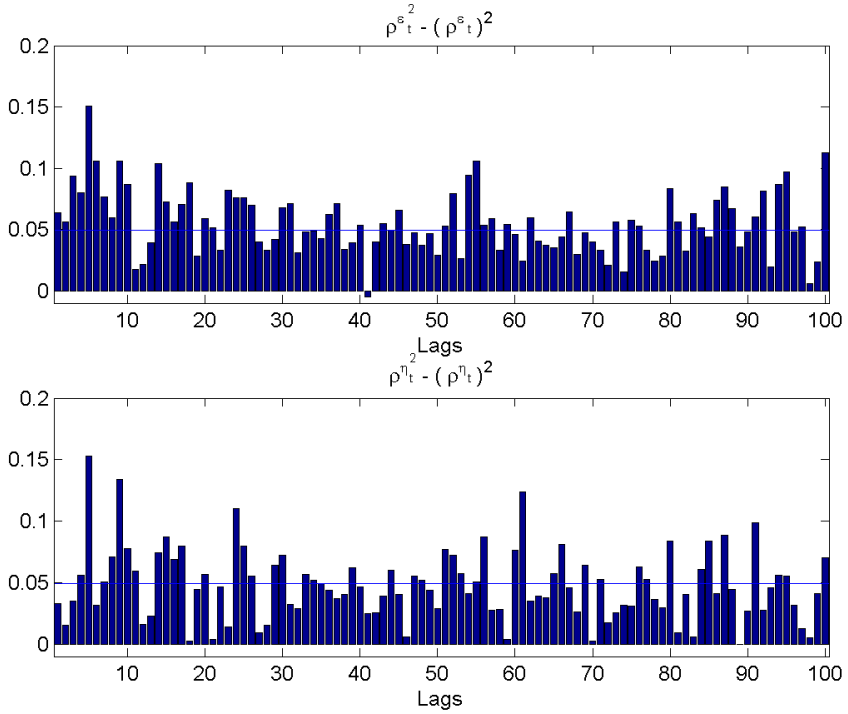

Figure 2.11: Correlogram of squares of the auxiliary residuals, $\hat{\varepsilon}_t$ and $\hat{\eta}_t$, corrected by the square of their sample autocorrelations.

component.

In reference to the IMA-GARCH model, we can see that the overall fit is almost identical to the LL-GARCH model. Furthermore, it is also clear that the nonlinear model improves the fit in relation with the homoscedastic IMA model. However, as stated before, the estimated ARCH coefficient, $\hat{\delta}_1$ is smaller than both, $\hat{\alpha}_1$ and $\hat{\gamma}_1$, suggesting that the level of heteroscedasticity of the reduced form model is inferior compared to both components. Furthermore, the value of $\hat{\delta}_1$ is roughly equal to the approximated value of $\delta_1 = 0.029$ obtained after plugging the estimates of the local level parameters into the system of equations given in (2.25), considered in Section 2.3.

2.5 Conclusions

In this chapter we derive the kurtosis and acf of squares of the reduced forms of two particular unobserved component models that are of empirical interest, the local level and smooth trend models. We particularize the expressions of these two moments to the cases in which the unobserved component disturbances are conditionally heteroscedastic in the form of either GARCH(1,1) or ARSV(1) processes. We also study the effects of fitting IMA-GARCH or IMA-ARSV models to conditionally heteroscedastic local level series.

We first show that if the noises of two the unobserved component models considered are mutually and serially uncorrelated with finite variances, the corresponding reduced form inno-

Local Level - GARCH(1,1)			IMA(1,1) - GARCH(1,1)		
	Estimates	(t-stat)		Estimates	(t-stat)
$\hat{\alpha}_0 =$	1.00E-04	(10.2)	$\hat{\delta}_0 =$	0.001	(9.18)
$\hat{\alpha}_1 =$	0.072	(73.6)	$\hat{\delta}_1 =$	0.026	(4.27)
$\hat{\alpha}_2 =$	0.920	(934.1)	$\hat{\delta}_2 =$	0.971	(158.6)
$\hat{\gamma}_0 =$	3.00E-04	(20.0)			
$\hat{\gamma}_1 =$	0.036	(81.8)			
$\hat{\gamma}_2 =$	0.961	(2023.8)			
	$\hat{v}_t^\dagger$				$\hat{a}_t^\dagger$
Mean	0.018			Mean	0.019
SK	0.144			SK	0.129
κ	3.350			κ	3.333
ρ_1	0.003			ρ_1	0.009
ρ_2	-0.033			ρ_2	-0.027
ρ_3	0.006			ρ_3	0.005
ρ_4	-0.012			ρ_4	-0.012
ρ_5	0.031			ρ_5	0.035
ρ_{10}	-0.025			ρ_{10}	-0.025
$Q(10)$	11.160			$Q(10)$	10.930
	$\hat{v}_t^{\dagger\,2}$				$\hat{a}_t^{\dagger\,2}$
ρ_1	0.001			ρ_1	0.006
ρ_2	0.011			ρ_2	0.006
ρ_3	0.009			ρ_3	0.013
ρ_4	0.004			ρ_4	0.002
ρ_5	0.073**			ρ_5	0.069**
ρ_{10}	0.001			ρ_{10}	0.001
$Q_1(10)$	18.78			$Q_1(10)$	18.27
$LogL$	-1372			$LogL$	-1374

Table 2.9: Estimates of the local level model with both noises being GARCH(1,1) and of the IMA(1,1)-GARCH(1,1) model, fitted to the £/€ exchange rate. The sample moments and correlograms reported refer to the residuals after being standardized by their estimated conditional variances. *(**) Significant at 5% (1%) level.

vation is also uncorrelated. Even more, if the noises of the local level are serially independent non-Gaussian processes, the reduced form noise is still uncorrelated although non-independent. On the other hand, we show that taking differences in series with conditionally heteroscedastic stochastic trends, weakens the strength of the heteroscedasticity. Consequently, when the sample size is small, often one cannot reject the null of homoscedasticity in series composed of one or more conditionally heteroscedastic components. Finally, we illustrate these results by fitting the two alternative models to the daily £/€ exchange rate.

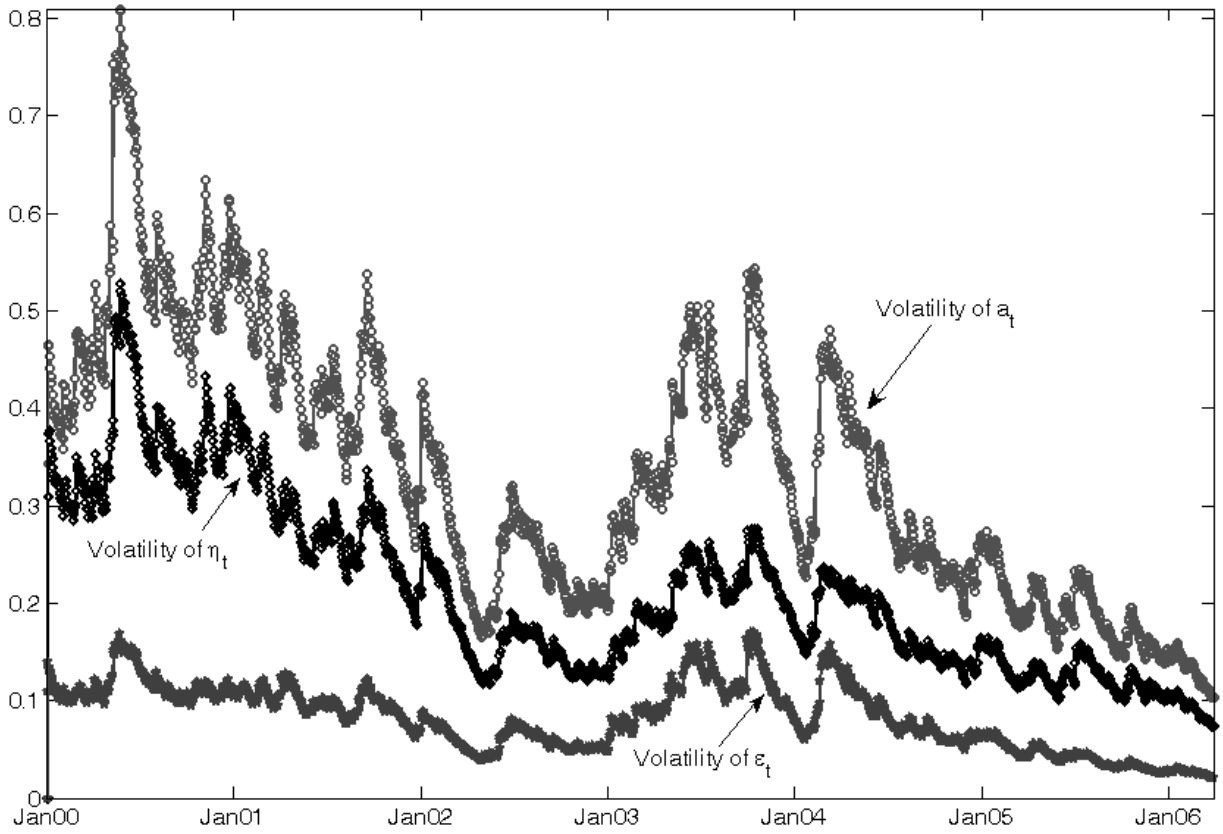

Figure 2.12: Estimated volatility of the daily £/€ exchange rate decomposed by its components.

Summarizing, in this chapter we show that although working with the ARIMA reduced form model is simpler because there is only one disturbance, working with the unobserved component model may lead to discover conditionally heteroscedastic structures that could not be apparent in the reduced form noise. In this regard, next chapter will be devoted to study the effects on prediction intervals of fitting ARIMA models to conditionally heteroscedastic series with stochastic levels.

Chapter 3

Prediction intervals of conditionally heteroscedastic unobserved component models

3.1 Introduction

In the previous chapter, we study the statistical properties of the reduced form of unobserved component models when they have conditionally heteroscedastic noises, and compare them to conditionally heteroscedastic ARIMA models. In this chapter, we analyze and compare the properties of prediction intervals constructed by each of these two alternative ways of modelling series with stochastic trends and conditionally heteroscedastic components. Again, we focus on the two models already considered in Chapter 2, namely, the local level and smooth trend models given by (1.2) and (1.4), respectively. Given that, as observed in the previous chapter, imposing either ARSV(1) or GARCH(1,1) processes in the unobserved component disturbances does not significantly affect the results, we only consider the cases in which the unobserved component disturbances are GARCH. For the reduced form ARIMA models, the evolution of uncertainty over time can be incorporated by assuming that a_t is a GARCH(1,1) process, as we also show in the previous chapter that the dependence in the second order moments of a_t can be well approximated by this model with a somehow weaker structure than these of the unobserved component noises.

In both alternatives, the amplitudes of the intervals change depending on whether the conditional variance at the moment of forecasting is larger or smaller than the marginal variance. Denoting by excess volatility, the difference between both variances, we show that in the unobserved component models, if only the transitory component is heteroscedastic the excess volatility disappears as the prediction horizon increases. That is, the prediction intervals obtained with the local level and smooth trend models converge to the intervals of the corresponding homoscedastic model. However, given that the reduced form IMA model always contains at least one unit root, the corresponding prediction intervals depend on the excess volatility for all prediction horizons. Consequently, when the excess volatility is positive (negative), the multi-

step prediction intervals based on the IMA models are too wide (too narrow), when compared with the intervals based on the corresponding unobserved component model. On the other hand, when the heteroscedasticity affects the long-run stochastic level, the prediction intervals constructed using both approaches are very similar.

The rest of this chapter is structured as follows. Section 3.2 derives the prediction intervals for the local level and smooth trend models with GARCH disturbances and for their corresponding IMA-GARCH models. Section 3.3 reports the results of several Monte Carlo experiments carried out to analyze the performance of the prediction intervals constructed from both approaches. Section 3.4 contains an empirical application. Finally, Section 3.5 concludes by summarizing the main findings.

3.2 Prediction intervals

In this section, we derive expressions of the prediction intervals obtained when the local level and smooth trend models with GARCH disturbances are fitted to represent the dynamic evolution of a heteroscedastic series with stochastic levels. Additionally, we also derive the intervals obtained when fitting the corresponding IMA-GARCH models. We will see that when the conditional heteroscedasticity affects only the transitory component there may be crucial differences between both prediction intervals.

3.2.1 The local level model

Consider first that y_t is given by the local level model in (1.2) with GARCH errors (henceforth LL-GARCH) as defined in (2.8) and that the objective is to forecast y_{T+k} given $\{y_1, y_2, ..., y_T\}$. It is well known that if the criterium is to minimize the mean squared forecast error (MSFE), then the optimal point predictor of y_{T+k}, denoted by $\hat{y}_{T+k}$, is its conditional mean, i.e. $\hat{y}_{T+k} = \underset{T}{E}(y_{T+k})$, where the T under the expectation means that it is conditional on the information available at time T. Note that from (1.2), it is easy to see that $\underset{T}{E}(y_{T+k}) = \underset{T}{E}(\mu_{T+k}) = \underset{T}{E}(\mu_T) \equiv \hat{\mu}_T$. Furthermore, using (2.8) it is possible to derive the variances of ε_{T+k} and η_{T+k} conditional on $\{y_1, y_2, ..., y_T\}$ as follows

$$\begin{aligned}
\underset{T}{E}(\varepsilon^2_{T+k}) &= \underset{T}{E}\left[(\varepsilon^{\dagger}_{T+k})^2\right] \underset{T}{E}(h_{T+k}) \\
&= \alpha_0 + (\alpha_1+\alpha_2)\,\underset{T}{E}(h_{T+k-1}) \\
&= \alpha_0\left[\frac{1-(\alpha_1+\alpha_2)^{k-1}}{1-\alpha_1-\alpha_2}\right] + (\alpha_1+\alpha_2)^{k-1}\,\underset{T}{E}(h_{T+1}) \\
&= \sigma^2_{\varepsilon} + (\alpha_1+\alpha_2)^{k-1}(\hat{h}_{T+1} - \sigma^2_{\varepsilon}), \quad k \geq 1, \qquad (3.1a)
\end{aligned}$$

where $\hat{h}_{T+1} = \underset{T}{E}(h_{T+1})$ and $\sigma^2_{\varepsilon} = \dfrac{\alpha_0}{1-\alpha_1-\alpha_2}$. By analogy, it is straightforward to show that

$$\underset{T}{E}(\eta^2_{T+k}) = \sigma^2_{\eta} + (\gamma_1+\gamma_2)^{k-1}(\hat{q}_{T+1} - \sigma^2_{\eta}), \quad k \geq 1, \qquad (3.1b)$$

where $\hat{q}_{T+1} = \underset{T}{E}(q_{T+1})$ and $\sigma_\eta^2 = \dfrac{\gamma_0}{1-\gamma_1-\gamma_2}$. Using the expressions given in (3.1), it is easy to show that the MSFE of $\hat{y}_{T+k}$ is given by

$$\begin{aligned} MSFE(\hat{y}_{T+k}) &= \underset{T}{E}\left[(y_{T+k} - \hat{y}_{T+k})^2\right] \\ &= \underset{T}{E}\left[(\mu_T + \eta_{T+1} + \ldots + \eta_{T+k} + \varepsilon_{T+k} - \hat{\mu}_T)^2\right] \\ &= P_T^\mu + \sigma_\varepsilon^2 + k\,\sigma_\eta^2 + \frac{1-(\gamma_1+\gamma_2)^k}{1-(\gamma_1+\gamma_2)}\left(\hat{q}_{T+1} - \sigma_\eta^2\right) \\ &\quad + (\alpha_1+\alpha_2)^{k-1}\left(\hat{h}_{T+1} - \sigma_\varepsilon^2\right),\ k = 1, 2, \ldots \end{aligned} \tag{3.2}$$

where $P_T^\mu = \underset{T}{E}[(\mu_T - \hat{\mu}_T)^2]$. Note that the LL-GARCH model is not conditionally Gaussian even when the standardized disturbances, $\varepsilon_t^\dagger$ and $\eta_t^\dagger$, are Gaussian. Therefore, obtaining the conditional expectations involved in the MSFE is not straightforward. Harvey et al. (1992) show that, in practice, it is possible to obtain approximations of the quantities $\hat{\mu}_T$, P_T^μ, $\hat{h}_{T+1}$ and $\hat{q}_{T+1}$ in (3.2) by using an augmented version of the Kalman Filter.

Expressions $(\hat{h}_{T+1} - \sigma_\varepsilon^2)$ and $(\hat{q}_{T+1} - \sigma_\eta^2)$ in (3.2) may be interpreted as measures of the excess volatility at the time of forecasting, with respect to the marginal variance in both noises. Note that the MSFE of the homoscedastic local level model is given by the first three terms of (3.2). Furthermore, given that $\alpha_1 + \alpha_2 < 1$, the MSFE of the LL-GARCH becomes a linear function of k in the long run, with the same slope as its homoscedastic counterpart, but with a different intercept due to the contribution of the fourth term in (3.2). However, for short and medium horizons, depending on whether the excess kurtosis is negative or positive, the influence of the excess volatility in both noises leads to a MSFE smaller or greater than that of the homoscedastic local level model.

Once $\hat{y}_{T+k}$ and its MSFE are available, one can obtain prediction intervals for y_{T+k} by assuming that the distribution of the k-steps-ahead prediction errors is Normal[1]. Therefore, the approximated $(1-\alpha)\%$ prediction intervals for the LL-GARCH model are given by

$$\hat{\mu}_T \ \pm\ Z_{\alpha/2}\sqrt{MSFE(\hat{y}_{T+k})}, \tag{3.3}$$

where $MSFE(\hat{y}_{T+k})$ is given by (3.2) and $Z_{\alpha/2}$ is the $\alpha/2$ quantile of the standard Normal density. On the other hand, when assuming homoscedasticity, the prediction intervals are given by

$$\hat{\mu}_T \ \pm\ Z_{\alpha/2}\sqrt{P_T^\mu + \sigma_\varepsilon^2 + k\,\sigma_\eta^2}. \tag{3.4}$$

It is important to note that there is a significant difference in the behavior of the prediction intervals in (3.3) depending on whether the conditional heteroscedasticity affects the long or the short-run components. An excess volatility in the permanent component affects the MSFE for all horizons while the effect of an excess volatility in the transitory component vanishes in the long run. Therefore, when the heteroscedasticity only affects the transitory noise, i.e.

[1] As we commented before, the prediction error distribution is not Gaussian. However, the results of Pascual et al. (2006) suggest that it could be well approximated by a Gaussian distribution.

$\gamma_1 = \gamma_2 = 0$, the prediction intervals in (3.3) converge to those of the homoscedastic model in (3.4). However, when the long run component is heteroscedastic, depending on the sign of the excess volatility, the prediction intervals of the heteroscedastic local level model are wider or thinner than those obtained in the homoscedastic model for all prediction horizons. As an illustration, Figure 3.1 plots the prediction intervals obtained for a series simulated by the local level model with parameters $\alpha_0 = 0.05$, $\alpha_1 = 0.10$, $\alpha_2 = 0.85$, $\gamma_1 = \gamma_2 = 0$ and $q_\eta = 1$, i.e. only ε_t is heteroscedastic. The time point when the prediction is made is selected in such a way that the excess volatility is positive. Assuming that the parameters are known and using the Kalman Filter proposed by Harvey et al. (1992) to approximate the MSFE, we construct the prediction intervals for the LL-GARCH (in solid lines) as in (3.3) and for the homoscedastic model (in dash-dotted lines) as in (3.4). Note that the LL-GARCH model produces wider intervals for short horizons than the homoscedastic model, because the conditional variance is higher than the marginal. However, since the shock producing the positive excess volatility is transitory, the prediction intervals of the LL-GARCH stick to those of the homoscedastic model as k increases. Figure 3.1 also plots possible trajectories of y_{T+k}, represented by the vertical clouds of points, in order to have a visual insight of the coverage of the prediction intervals for this particular series.

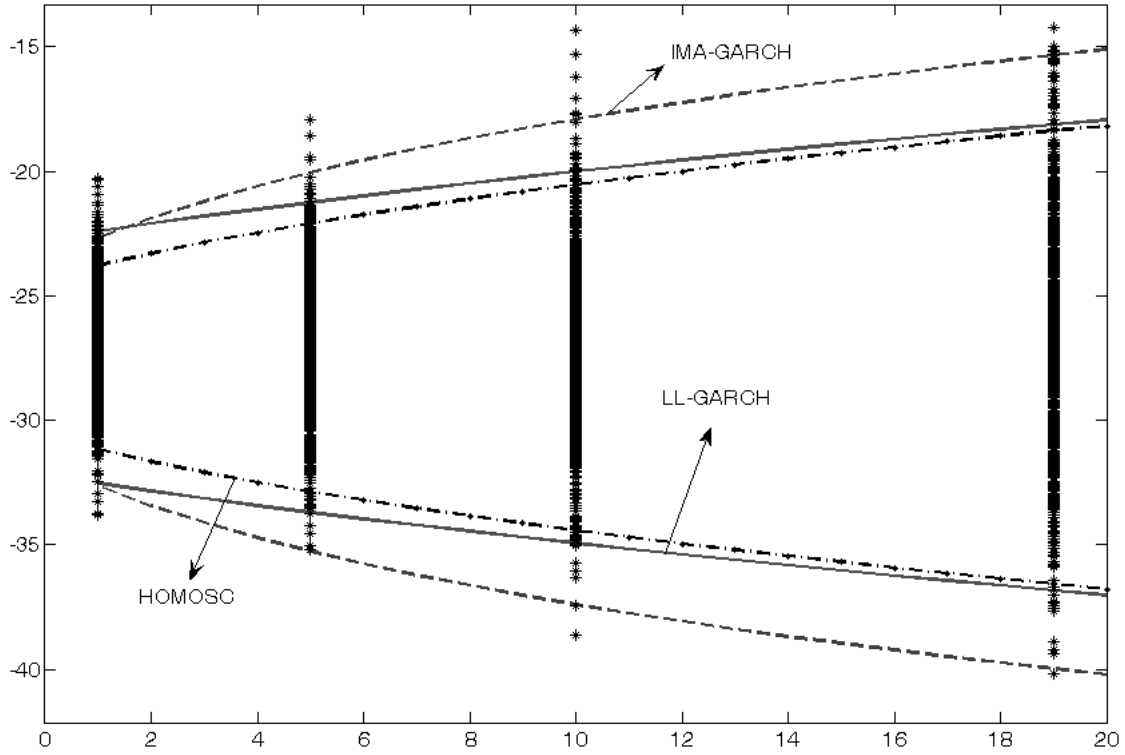

Figure 3.1: 95% Prediction intervals for a series simulated from the local level model with the transitory component being GARCH(1,1), with $\alpha_0 = 0.05$, $\alpha_1 = 0.10$, $\alpha_2 = 0.85$, and $q_\eta = 1$. The time point when the forecast is made is selected in a highly volatile period. The vertical clouds of points represent possible trajectories of y_{T+k}, given information at time T.

Consider now that instead of using the LL-GARCH, the corresponding IMA(1,1)-GARCH(1,1) model given by (2.1) and (2.28) is fitted to the series of interest. In this case, assuming that the

within sample innovations are observable, the optimal predictor of y_{T+k} given the information available at time T is given by

$$\hat{y}_{T+k} = y_T + \theta a_T, \ k = 1, 2, ... \tag{3.5}$$

with MSFE

$$MSFE(\hat{y}_{T+k}) = \begin{cases} \sigma_a^2 + (\sigma_{T+1}^2 - \sigma_a^2), & k = 1 \\ \left[(1+\theta)^2(k-1)+1\right]\sigma_a^2 + & \\ \left[\frac{(1+\theta)^2 - (\delta_1+\delta_2)^{k-1}(\theta(2+\theta)+\delta_1+\delta_2)}{1-(\delta_1+\delta_2)}\right](\sigma_{T+1}^2 - \sigma_a^2), & k = 2, 3, ... \end{cases} \tag{3.6}$$

Once more, $(\sigma_{T+1}^2 - \sigma_a^2)$ is a measure of the excess volatility of a_t. Note that the MSFE in (3.6) can also be separated into a linear and a nonlinear part, defined by the first and second terms, respectively. It is clear from (3.6) that as k increases, the $MSFE(\hat{y}_{T+k})$ is also a linear function of the horizon. However, as long as the excess volatility is different from zero, the path of MSFE in the IMA-GARCH model is always above or below the path of the MSFE in the corresponding homoscedastic IMA model. This implies that the sign of the excess volatility at time T determines if the IMA-GARCH prediction variance will be smaller or larger than the prediction variance of the homoscedastic IMA for all prediction horizons. In this sense, the behavior is similar to that of the local level model with heteroscedastic long-run disturbances.

As in the LL-GARCH model, k-steps-ahead intervals based on the IMA-GARCH model can be obtained by assuming that the forecast errors are Normally distributed for all k. In this case, the approximated $(1-\alpha)\%$ prediction intervals for the IMA-GARCH model are given by

$$y_T + \theta\, a_T \ \pm \ Z_{\alpha/2}\sqrt{MSFE(\hat{y}_{T+k})}, \tag{3.7}$$

where the $MSFE(\hat{y}_{T+k})$ is given by (3.6). Finally, we can construct the prediction intervals for future values of y_{T+k} by assuming an homoscedastic IMA model. We do not further consider these intervals as they are identical to the ones obtained using the homoscedastic local level model in (3.4).

Going back to the illustration in Figure 3.1, we also plot the IMA-GARCH prediction intervals (in dashed lines) from (3.7). The values of θ and δ_0 in the IMA-GARCH model have been obtained by using the functions that relate them with the signal-to noise ratio, q_η, and the marginal variance, σ_ε^2, while the parameters of the GARCH process, δ_1 and δ_2, have been recovered from α_1 and α_2 following the procedure described in Chapter 2. By looking at the resulting intervals, we observe that they have almost the same length as those of the LL-GARCH for very short horizons, but they become wider as k increases. This behavior is the consequence of taking the transitory shock as permanent, so that the positive excess volatility leads to a higher MSFE and wider prediction intervals.

3.2.2 The smooth trend model

Consider now that y_t is given by the smooth trend model in (1.4), with ε_t being a GARCH(1,1) process (henceforth ST-GARCH). In this case, the optimal point predictor is given by

$$\hat{y}_{T+k} = \hat{\mu}_T + k\,\hat{\beta}_T, \ k = 1, 2, ... \tag{3.8}$$

where $\hat{\beta}_T = \underset{T}{E}(\beta_T)$. Note that, in this case, the optimal point predictor grows linearly with the time horizon, k. From (3.8) we find the following expression of the MSFE when ε_t is GARCH,

$$\begin{aligned} MSFE(\hat{y}_{T+k}) &= P_T^{\mu} + k^2\, P_T^{\beta} + 2k\, P_T^{\mu,\beta} + \frac{k(k-1)(2k-1)}{6}\sigma_\xi^2 + \underset{T}{E}[\varepsilon_{T+k}^2] \\ &= P_T^{\mu} + k^2\, P_T^{\beta} + 2k\, P_T^{\mu,\beta} + \frac{k(k-1)(2k-1)}{6}\sigma_\xi^2 + \sigma_\varepsilon^2 \\ &\quad + (\alpha_1+\alpha_2)^{k-1}\,(\hat{h}_{T+1} - \sigma_\varepsilon^2),\ k = 1, 2, ..., \end{aligned} \tag{3.9}$$

where $P_T^{\beta} = \underset{T}{E}[(\beta_T - \hat{\beta}_T)^2]$ and $P_T^{\mu,\beta} = \underset{T}{E}[(\mu_T - \hat{\mu}_T)(\beta_T - \hat{\beta}_T)]$. Note that the MSFE of the homoscedastic smooth trend model is given by the first five terms of (3.9), and the last term depends on the excess volatility. However, the long run predictions of the homoscedastic and GARCH smooth trend model will have the same MSFEs since the heteroscedasticity comes only from the transitory component. Following the same arguments of Harvey et al. (1992) for the LL-GARCH model, it is possible to obtain estimates of the conditional moments involved in (3.8) and (3.9) by using an augmented version of the Kalman Filter. Then, it is also possible to construct approximated $(1-\alpha)\%$ prediction intervals for the ST-GARCH model as follows,

$$\hat{\mu}_T + k\,\hat{\beta}_T \quad \pm \quad Z_{\alpha/2}\,\sqrt{MSFE(\hat{y}_{T+k})}, \tag{3.10}$$

where $MSFE(\hat{y}_{T+k})$ is given by (3.9). On the other hand, the homoscedastic prediction interval are constructed as follows,

$$\hat{\mu}_T + k\,\hat{\beta}_T \quad \pm \quad Z_{\alpha/2}\,\sqrt{P_T^{\mu} + k^2\, P_T^{\beta} + 2k\, P_T^{\mu,\beta} + \frac{k(k-1)(2k-1)}{6}\sigma_\xi^2 + \sigma_\varepsilon^2}. \tag{3.11}$$

As an illustration, Figure 3.2 presents a case in which the prediction intervals of simulated series are calculated at a highly volatile period. For the simulations, we set $\sigma_\varepsilon^2 = 1$, $\alpha_1 = 0.15$, $\alpha_2 = 0.8$, and $\sigma_\xi^2 = 0.5$ (i.e $q_\xi = 0.5$).

Again, if we are now interested in fitting the IMA(2,2)-GARCH(1,1) of (2.2) and (2.28), then the optimal predictor is given by

$$\hat{y}_{T+k} = \begin{cases} 2y_T - y_{T-1} + \theta_1\, a_T + \theta_2\, a_{T-1}, & k = 1 \\ k\,\hat{y}_{T+1} - (k-1)(y_T - \theta_2\, a_T), & k = 2, 3, ... \end{cases} \tag{3.12}$$

with MSFE

$$MSFE(\hat{y}_{T+k}) = \begin{cases} \sigma_a^2 + (\sigma_{T+1}^2 - \sigma_a^2), & k = 1 \\ (k-1)\left[(1-\theta_2)^2 + \frac{k\,(2k-1)+6(1-\theta_2)\,k}{6}(1+\theta_1+\theta_2)^2\right]\sigma_a^2 & \\ + \sum_{j=0}^{k-1} B_j^2\,(\delta_1+\delta_2)^{k-1-j}\ (\sigma_{T+1}^2 - \sigma_a^2), & k = 2, 3, ... \end{cases} \tag{3.13}$$

where $B_0 = 1$ and $B_j = j\,(1+\theta_1+\theta_2)+(1-\theta_2)$, $j = 1, 2, ..., k-1$. In this case, the first term is the linear part of the MSFE and the second the nonlinear one, defined as a function of the excess volatility and the prediction horizon, k. When compared to the MSFE of the smooth trend model, the nonlinear term does not vanish when k increases. Therefore, as in the IMA(1,1)-GARCH case, any shock that leads to an excess volatility different from zero is permanent and

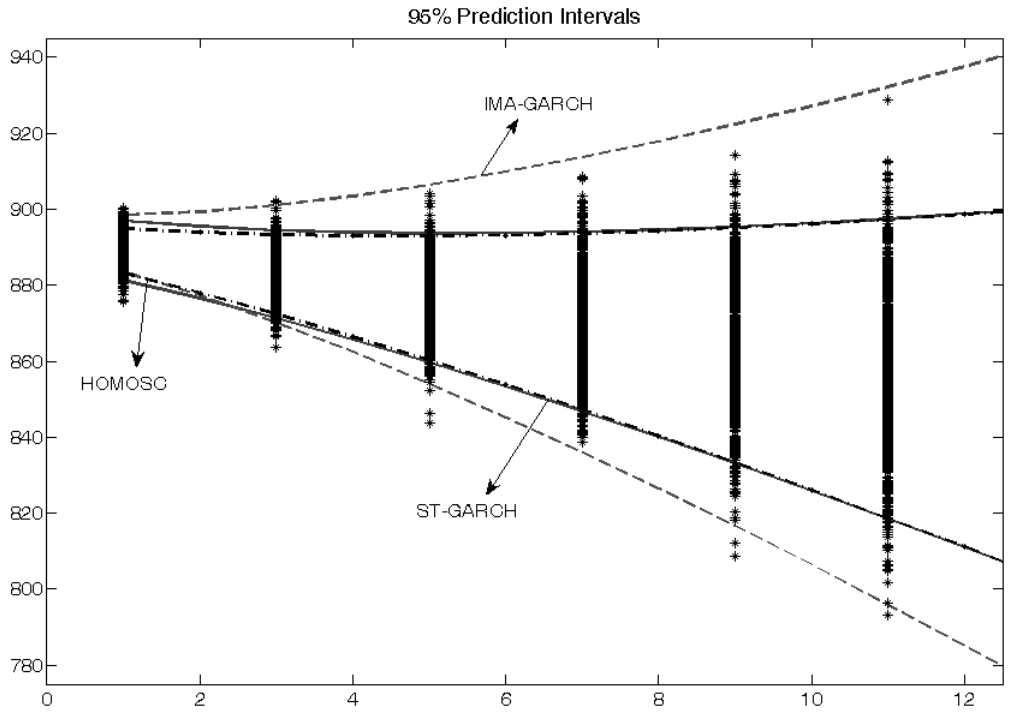

Figure 3.2: Prediction intervals of an ST-GARCH series, with $\alpha_1 = 0.15$, $\alpha_2 = 0.8$, and $q_\xi = 0.5$. The time point is selected in a highly volatile period. The vertical clouds of points represent possible trajectories of y_{T+k}, given information at time T.

thus the MSFE of the heteroscedastic model will differ from that of the homoscedastic one for all prediction horizons. By analogy to the IMA-GARCH case in the local level model, the prediction intervals constructed from (3.13) are given by

$$\hat{y}_{T+k} \quad \pm \quad Z_{\alpha/2}\sqrt{MSFE(\hat{y}_{T+k})}, \tag{3.14}$$

with $\hat{y}_{T+k}$ and $MSFE(\hat{y}_{T+k})$ defined as in (3.12) and (3.13), respectively. Figure 3.2 shows the IMA-GARCH prediction intervals in dotted lines. Again, as in the local level case, it seems that the IMA-GARCH prediction intervals are too wide in terms of the nominal coverage. Indeed, they contain almost all observations for all k.

Summarizing, in the models where the transitory component, ε_t, is the only heteroscedastic component, the shocks to the variance are purely transitory. Consequently, the homoscedastic and the unobserved component with GARCH noises produce prediction intervals that stick to each other as k increases. However, depending on the sign of the excess volatility, the ARIMA-GARCH counterparts may be wider or thinner than the intervals obtained with the corresponding unobserved component model. This is due to its incapacity of distinguishing whether the heteroscedasticity affects the long or the short run components, and it may lead to significant differences between the two prediction intervals, specially for medium and long term. In the next section we will study this issue in depth by performing several Monte Carlo experiments and reporting the nominal coverage of each prediction intervals.

3.3 Forecasting Performance

Having seen that the conclusions for the LL-GARCH model can be applicable for the ST-GARCH model, in this section we will focus on the former one. In order to analyze the performance of the prediction intervals described in the previous section, we calculate the observed coverage of these intervals by generating $B = 1000$ trajectories of y_{T+k} conditional on $\{y_1, y_2, ..., y_T\}$. Then, we repeat this procedure for 1000 series simulated from the LL-GARCH. As an example, Figure 3.3 reports the observed coverage of the 95% prediction intervals shown in Figure 3.1. Note that the coverage of the three prediction intervals is different across models and across the forecast horizons k.

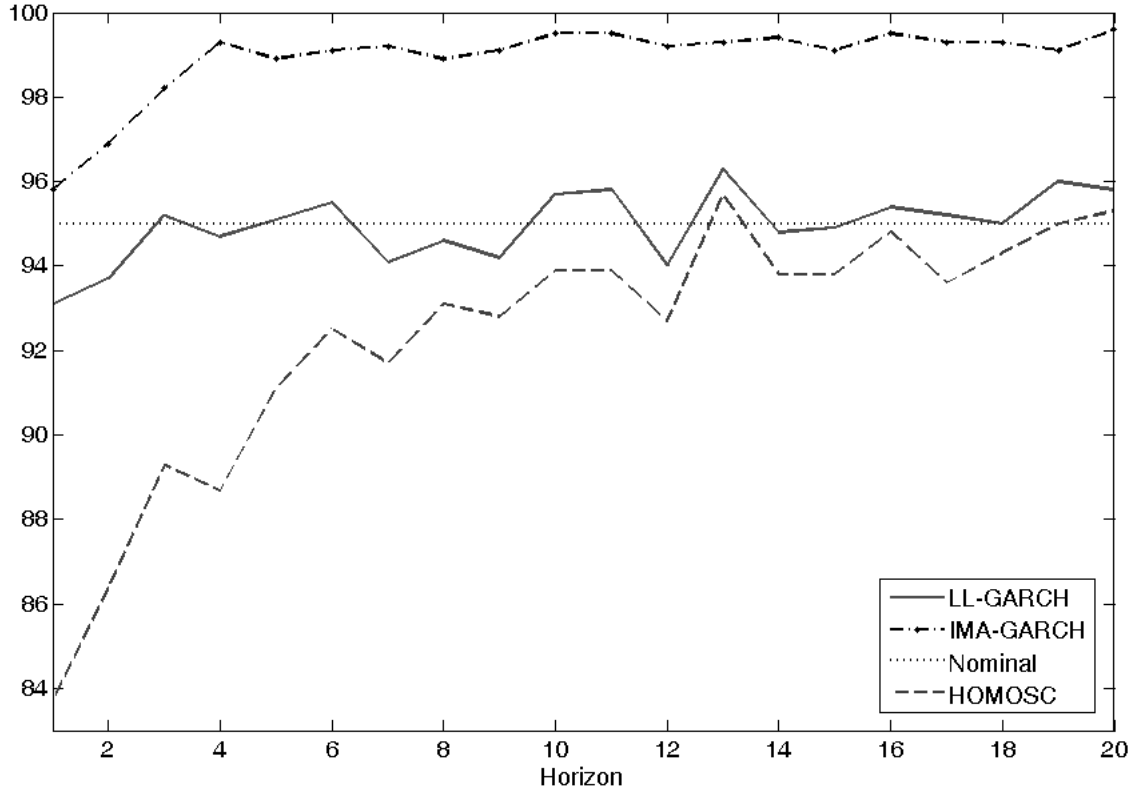

Figure 3.3: Observed coverage measured as the percentage of trajectories within the 95% prediction intervals of the three models selected. The series is simulated from the local level model with the transitory component being GARCH, and the time point corresponds to a highly volatile period; see Figure 3.1.

For the Monte Carlo simulations, we consider four designs depending on whether the transitory or the permanent components are heteroscedastic, and on the value of the signal-to-noise ratio. For each design, we construct 90% and 95% prediction intervals from the estimated LL-GARCH model using the augmented Kalman Filter, the homoscedastic LL model and the IMA-GARCH model as given in (3.3), (3.4) and (3.7), respectively. In the first two designs, the transitory component is heteroscedastic while the long-run component is homoscedastic. Their parameters are $\alpha_0 = 0.05$, $\alpha_1 = 0.10$ and $\alpha_2 = 0.85$ with $q_\eta = 0.5$ in the first model and $q_\eta = 1$ in the second. In the other two designs, the long-run component is heteroscedastic with $\gamma_0 = 0.05$, $\gamma_1 = 0.10$, $\gamma_2 = 0.85$ and $q_\eta = 1$ in the third model and $q_\eta = 2$ in the last one. The prediction

horizons considered are $k = 1,\ 2,\ 6,\ 12$ and 24.

Note that, in practice, the parameters needed to construct the prediction intervals in (3.3), (3.4) and (3.7) should be estimated. Consequently, in order to compare the effects of estimation on the performance of prediction intervals, we construct them by assuming known parameters and by substituting the parameters by their QML estimates. To estimate the parameters of the heteroscedastic local level model, we follow the estimation approach proposed by Harvey et al. (1992). Tables 3.1 and 3.2 report the mean absolute deviation (MAD) of the observed coverages with respect to the nominal, for the two models with heteroscedasticity in the transitory component and the two models with heteroscedastic long-run component, respectively. The first conclusion from both tables is that regardless of the model and coverage considered, the deviations between the empirical and nominal coverages are in general smaller as the prediction horizon increases. However, the conclusions arising from Tables 3.1 and 3.2 are rather different.

Consider first, the results reported in Table 3.1 for the two models in which the long-run component is homoscedastic. In this case, if the parameters are known, we can observe that the homoscedastic models have the largest deviations between empirical and nominal coverages only when predicting in the short run for $k = 1$. However, when the prediction horizon is larger, constructing the intervals using the IMA model leads to the worst coverages. This is due to the inability of the IMA-GARCH model to represent the dynamics of series in which only the transitory component is heteroscedastic. Also note that when the forecast horizon is large ($k = 24$), the deviations between the empirical and nominal coverages of the homoscedastic and LL-GARCH models are similar. Obviously, when the parameters are estimated, we observe larger deviations. However, the conclusions about the comparison among intervals are the same as those obtained assuming known parameters.

We consider now the results reported in Table 3.2 for the models in which the short-run component is homoscedastic. In this case, regardless of whether the parameters are known or estimated, the homoscedastic prediction intervals are worse than any of the two intervals constructed with the heteroscedastic models for all prediction horizons. Furthermore, comparing the heteroscedastic intervals between them, we observe that their MAD are very similar, mainly when the parameters are estimated. Therefore, when the long-run component is heteroscedastic, constructing the prediction intervals using the LL-GARCH model, only leads to slight improvements in the performance of prediction intervals with respect to constructing them using the simpler IMA-GARCH models. Once more, the deviations between empirical and nominal coverages are larger when the parameters are estimated.

	90%				95%			
Parameters: $\alpha_1 = 0.10$, $\alpha_2 = 0.85$, $q = 1$	$k = 1$	$k = 6$	$k = 12$	$k = 24$	$k = 1$	$k = 6$	$k = 12$	$k = 24$
Homoscedastic (true param)	3.164	1.210	0.837	0.784	2.186	0.867	0.633	0.560
Augmented Kalman Filter (true param)	1.575	0.885	0.742	0.773	1.104	0.630	0.571	0.549
IMA-GARCH(1,1) (true param)	1.753	1.823	1.877	1.684	1.220	1.249	1.297	1.173
Homoscedastic (QML estim)	3.241	1.593	1.463	1.578	2.260	1.134	1.022	1.064
Augmented Kalman Filter (QML estim)	2.011	1.442	1.515	1.531	1.402	1.000	1.077	1.043
IMA-GARCH(1,1) (QML estim)	2.179	2.079	2.176	2.104	1.483	1.387	1.491	1.409
Parameters: $\alpha_1 = 0.10$, $\alpha_2 = 0.85$, $q = 0.5$								
Homoscedastic (true param)	3.906	1.685	1.098	0.825	2.705	1.164	0.774	0.583
Augmented Kalman Filter (true param)	1.792	0.959	0.806	0.764	1.223	0.703	0.583	0.546
IMA-GARCH(1,1) (true param)	1.930	1.933	2.019	1.934	1.306	1.332	1.392	1.280
Homoscedastic (QML estim)	3.940	2.034	1.735	1.683	2.698	1.395	1.210	1.179
Augmented Kalman Filter (QML estim)	2.064	1.505	1.541	1.635	1.442	1.067	1.088	1.134
IMA-GARCH(1,1) (QML estim)	2.206	2.110	2.288	2.338	1.534	1.422	1.536	1.550

Table 3.1: Mean Absolute Deviation (MAD) of the differences between observed and nominal coverages for horizons, $k = 1,\ 6,\ 12$ and 24, and two confidence levels, 90% and 95%. The MAD is calculated in percentages. The series are simulated from the local level model with a GARCH(1,1) process in the transitory component, ε_t.

	90%				95%			
Parameters: $\gamma_1 = 0.10$, $\gamma_2 = 0.85$, $q = 1$	$k = 1$	$k = 6$	$k = 12$	$k = 24$	$k = 1$	$k = 6$	$k = 12$	$k = 24$
Homoscedastic (true param)	2.325	3.606	3.411	2.804	1.607	2.421	2.310	1.850
Augmented Kalman Filter (true param)	1.624	2.631	2.497	2.087	1.132	1.775	1.711	1.400
IMA-GARCH(1,1) (true param)	1.715	2.933	2.813	2.384	1.193	1.946	1.905	1.552
Homoscedastic (QML estim)	2.387	3.723	3.656	3.236	1.659	2.570	2.5210	2.216
Augmented Kalman Filter (QML estim)	1.858	3.000	2.985	2.727	1.299	2.095	2.098	1.945
IMA-GARCH(1,1) (QML estim)	1.976	3.144	3.139	2.804	1.384	2.195	2.220	2.013
Parameters: $\gamma_1 = 0.10$, $\gamma_2 = 0.85$, $q = 2$								
Homoscedastic (true param)	2.995	3.755	3.471	2.808	2.060	2.527	2.262	1.868
Augmented Kalman Filter (true param)	1.969	2.483	2.299	1.904	1.355	1.698	1.524	1.286
IMA-GARCH(1,1) (true param)	2.016	2.715	2.542	2.119	1.384	1.833	1.668	1.421
Homoscedastic (QML estim)	3.077	3.972	3.668	3.159	2.124	2.689	2.489	2.142
Augmented Kalman Filter (QML estim)	2.241	2.933	2.811	2.612	1.551	2.040	1.976	1.833
IMA-GARCH(1,1) (QML estim)	2.320	3.077	2.974	2.648	1.600	2.165	2.091	1.918

Table 3.2: Mean Absolute Deviation (MAD) of the differences between observed and nominal coverages for horizons, $k = 1,\ 6,\ 12$ and 24, and two confidence levels, 90% and 95%. The MAD is calculated in percentages. The series are simulated from the local level model with a GARCH(1,1) process in the permanent component, η_t.

The information in Tables 3.1 and 3.2 allows us to compare the alternative models in terms of which generates smallest deviations between the nominal and empirical coverages. However, it does not contain any information on the sign of these deviations. Therefore, we cannot conclude whether we are obtaining intervals that cover more or less than the nominal. Obviously, the sign of the deviations depends on the sign of the excess volatility at the time the prediction is made. Consequently, we also compute the mean coverage of each model and prediction horizon, conditional on whether the excess volatility is positive or negative. Figure 3.4 plots the averaged empirical coverages against the horizon k, computed in a volatile period, when the marginal variance is smaller than the conditional (top row) and in a quiet period in which the marginal is larger than the conditional (bottom row) for the model in which the short-run component is heteroscedastic and $q_\eta = 0.5$. Figure 3.4 shows that in both periods, the coverage of the homoscedastic model tends to the nominal when the prediction horizon increases. However, in the short-run, the coverage of the homoscedastic prediction intervals is smaller (larger) than the nominal when the excess volatility is positive (negative). The short-run coverages of the two heteroscedastic intervals are better than those of the homoscedastic intervals, although there still exists a gap between the empirical and nominal coverages.[2]. On the other hand, the long-run coverages of the IMA-GARCH intervals are well above (below) the nominal when the excess volatility is positive (negative). As we mentioned above, this model incorporates the unit root and it cannot cope with the fact that only the transitory component is heteroscedastic. Finally, the average coverages of the LL-GARCH intervals are close to the nominal for all the prediction horizons considered.

Finally, note in Figure 3.4 that obtaining the prediction intervals with estimated parameters imply larger deviations of the empirical with respect to the nominal coverages. However, the conclusions on the comparative performance of the three intervals considered in this paper are the same. The only point worth to be made in this case is that estimating the parameters renders IMA-GARCH intervals with closer (further) coverages from the nominal when the period of forecasting is quiet (volatile) than those obtained when the parameters are known. This is due to the fact that when the prediction is made in a quiet period, the IMA-GARCH intervals tend to be too thin in order to have the assumed coverage. However, estimating the parameters generate wider intervals, which obviously have coverages closer to the nominal. On the other hand, when forecasting in a volatile period, estimating the parameters still yields wider intervals, which produce coverages even further from the nominal than when the parameters are known.

3.4 An empirical illustration

In this section, we construct and evaluate out-of-sample prediction intervals of two real time series. The series selected are the seasonally adjusted monthly US inflation rate and the daily Pound/Euro (£/€) exchange rate already described in Chapter 2. The inflation rate is defined as

[2]This happens presumably because the prediction intervals are based on quantities computed by running the Kalman filter as if the model were conditionally Gaussian while, in fact, the series are generated from the LL-GARCH model, which is not conditionally Gaussian.

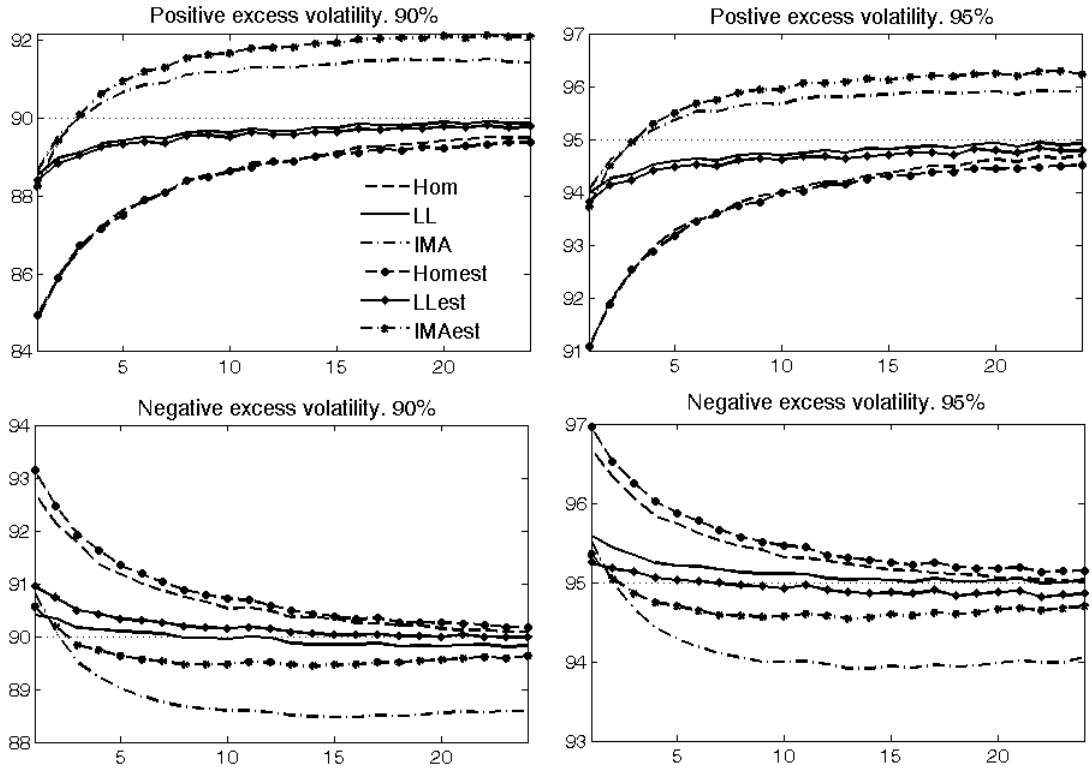

Figure 3.4: Mean observed coverage in volatile (positive excess volatility) and quiet (negative excess volatility) periods. The series are generated from the local level model with a GARCH(1,1) noise only in the transitory component. The parameters are given by $\alpha_1 = 0.10$, $\alpha_2 = 0.85$ and $q_\eta = 0.5$.

the log-difference of the monthly personal consumption expenditure (PCE) deflator, multiplied by 100 to have percentage rates, and it has been downloaded from the EcoWin database. As done in the last chapter, an intervention analysis prior to estimation was carried out to account for possible outliers. The period analyzed for this series spans 49 years, from February 1959 to May 2008, thus containing $T = 592$ observations. On the other hand, the exchange rate series is defined as 100 times the logarithm of the daily closing price of one Pound per unit of Euro. For this series, we use a different sample compared with that used in Chapter 2, covering from 1993 to 2003, with $T = 2589$ observations. We decide to work with this sample to evaluate the forecasting performance of the series immediately after the beginning of the Euro as a common currency. Figure 3.5 plots the two series, together with their corresponding sample autocorrelations of the observations in levels and in squares. Having shown in the last chapter that the local level and IMA(1,1) models provide a good fit to the exchange rate series, the correlogram of the first difference of the inflation rate seems to support the use of the same models for this series.

We proceed as follows. We use the first $T - R$ observations for the in-sample estimation of the three models, namely, the homoscedastic local level, the LL-GARCH and IMA-GARCH models, and construct the prediction intervals k steps ahead, conditional on the estimated parameters. Then, we move the sample window one observation at a time, re-estimate the models and construct again the prediction intervals conditional on the new estimates. At the end of the procedure, we have $R - k$ k-steps-ahead prediction intervals for each model and each

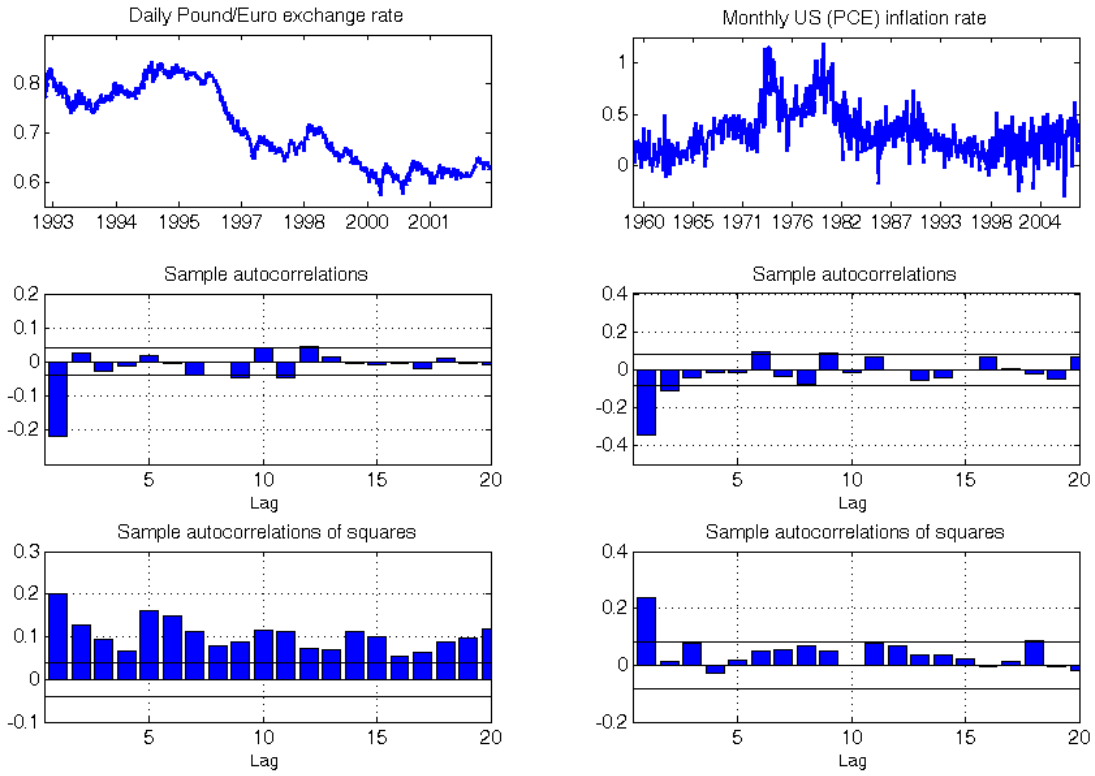

Figure 3.5: The daily £/€ exchange rate, from 1993 to 2003, and the monthly seasonally adjusted US inflation rate, from 1959 to 2008.

series. With them, we find the out-of-sample observed coverage, defined as the proportion of actual observations laying within the intervals. We fix $R = 450$ in the exchange rate series (from January 2000 to September 2003) and $R = 90$ in the inflation rate series (from Dec 2000 to May 2008). In the inflation series, given that the variance of the permanent component is very small relative to that of the transitory component (i.e. the estimated $\hat{q}_\eta$ is very small), we assume that the permanent component noise, η_t, is homoscedastic and include GARCH(1,1) effects only in the transitory component, ε_t. On the other hand, we include GARCH effects in both noises of the local level model fitted to the exchange rate series.

Table 3.3 shows the in-sample estimation results, with the first $T-R$ observations. Note that the estimated signal-to-noise ratio for the exchange rate series is much larger than that of the inflation series. This implies that the volatility shocks in the permanent (transitory) component will have a great impact in the exchange rate (inflation rate) series. Figure 3.6 plots the series of time-varying volatility throughout the sample, in order to assess the size of the excess volatilities in each component with respect to its marginal variance, represented by the horizontal dashed lines. Note that, in the exchange rate series, the volatility in the transitory component, $\hat{h}_t$, is important only in the first part of the sample. Consequently, in the evaluation period, after the vertical dotted line, all volatility shocks are driven by the permanent component. On the other hand, the transitory component in the inflation series is the only heteroscedastic one. In other words, $(\hat{q}_t - \sigma_\eta^2) = 0$ for all t, so that the MSFE of the LL-GARCH model will differ from that of the homoscedastic model only when $(\hat{h}_t - \sigma_\varepsilon^2) \neq 0$. As we can see from the plot of $\hat{h}_t$, this

difference is positive in almost all months of the evaluation period.

Pound/Euro exchange rate				US inflation rate			
LL-GARCH		IMA-GARCH		LL-GARCH		IMA-GARCH	
$\hat{\sigma}^2_\varepsilon =$	8.3×10^{-2}	$\hat{\sigma}^2_a =$	34.1×10^{-2}	$\hat{\sigma}^2_\varepsilon =$	16.1×10^{-3}	$\hat{\sigma}^2_a =$	21.7×10^{-3}
$\hat{\sigma}^2_\eta =$	19.6×10^{-2}	$\hat{\theta} =$	-0.243**	$\hat{\sigma}^2_\eta =$	14.9×10^{-3}	$\hat{\theta} =$	-0.738**
$\hat{\alpha}_1 =$	0.131**	$\hat{\delta}_1 =$	0.049**	$\hat{\alpha}_1 =$	0.193**	$\hat{\delta}_1 =$	0.141**
$\hat{\alpha}_2 =$	0.866**	$\hat{\delta}_2 =$	0.942**	$\hat{\alpha}_2 =$	0.738**	$\hat{\delta}_2 =$	0.803**
$\hat{\gamma}_1 =$	0.042**			$\hat{\gamma}_1 =$	–		
$\hat{\gamma}_2 =$	0.948**			$\hat{\gamma}_2 =$	–		

Table 3.3: Estimates of the LL-GARCH and IMA-GARCH models fitted to the £/€ exchange rate and the US inflation rate. *(**) Significant at 5% (1%) level.

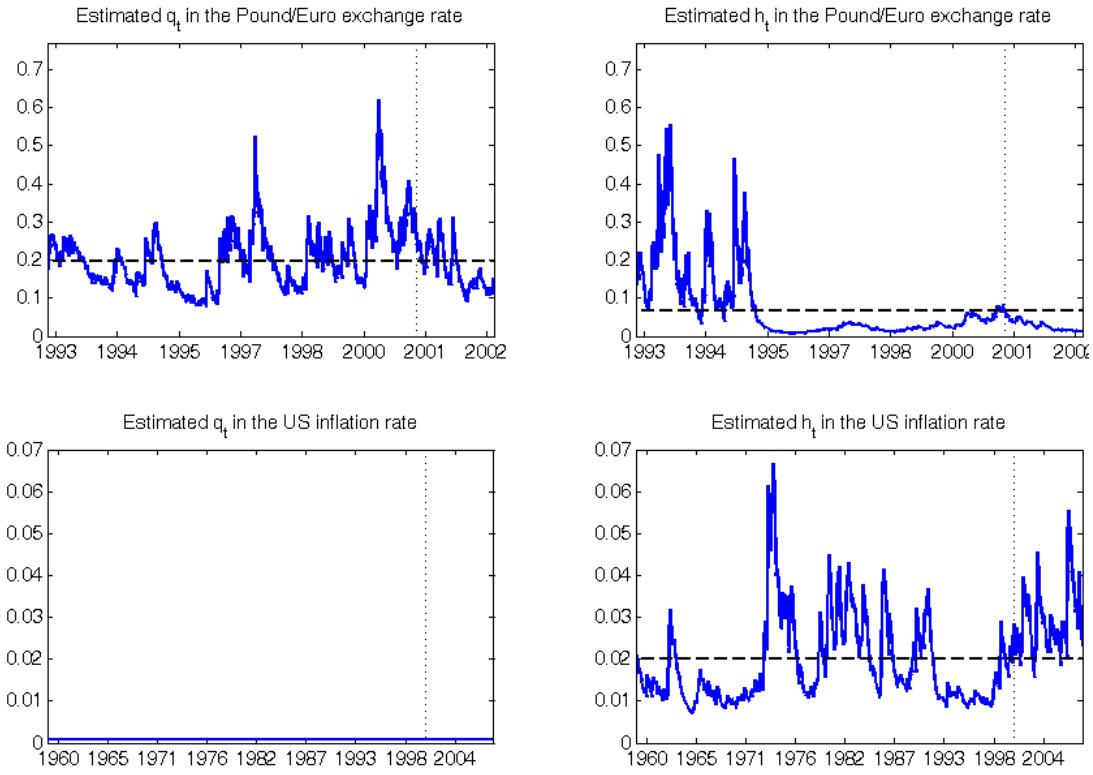

Figure 3.6: Conditional volatilities of £/€ exchange rate and the US inflation rate series (in solid lines), estimated from the local level with GARCH noises. h_t and q_t are the conditional variances of the transitory and permanent components, respectively. The estimated marginal variance of each component is represented by the horizontal dashed line. Finally, the vertical dotted line indicates the starting point of the evaluation period.

Conditional on the estimated parameters of each model, we find the approximated $(1-\alpha)\%$ prediction intervals for horizons running from $k = 1$ to $k = 36$ of the two heteroscedastic and the homoscedastic models given in (3.3), (3.4) and (3.7). We consider $\alpha = 5\%$ and 10%. As an illustration, Figure 3.7 plots the approximated 90% prediction intervals for the US inflation rate in June 2003, a period of high volatility, i.e. $(\hat{h}_t - \sigma^2_\varepsilon) > 0$. Observe that the homoscedastic model

produces much narrower prediction intervals than those produced by the two heteroscedastic models, LL-GARCH and IMA-GARCH, for small k. However, as the horizon increases, the LL-GARCH prediction intervals stick to those of the homoscedastic local level model, while those of the IMA-GARCH become wider. As we have seen with the simulated data, this behavior is explained by that fact that the IMA-GARCH takes a transitory shock as permanent. In this case, the shock is positive, so that the IMA-GARCH prediction intervals are always wider than those of the LL-GARCH when k is relatively large.

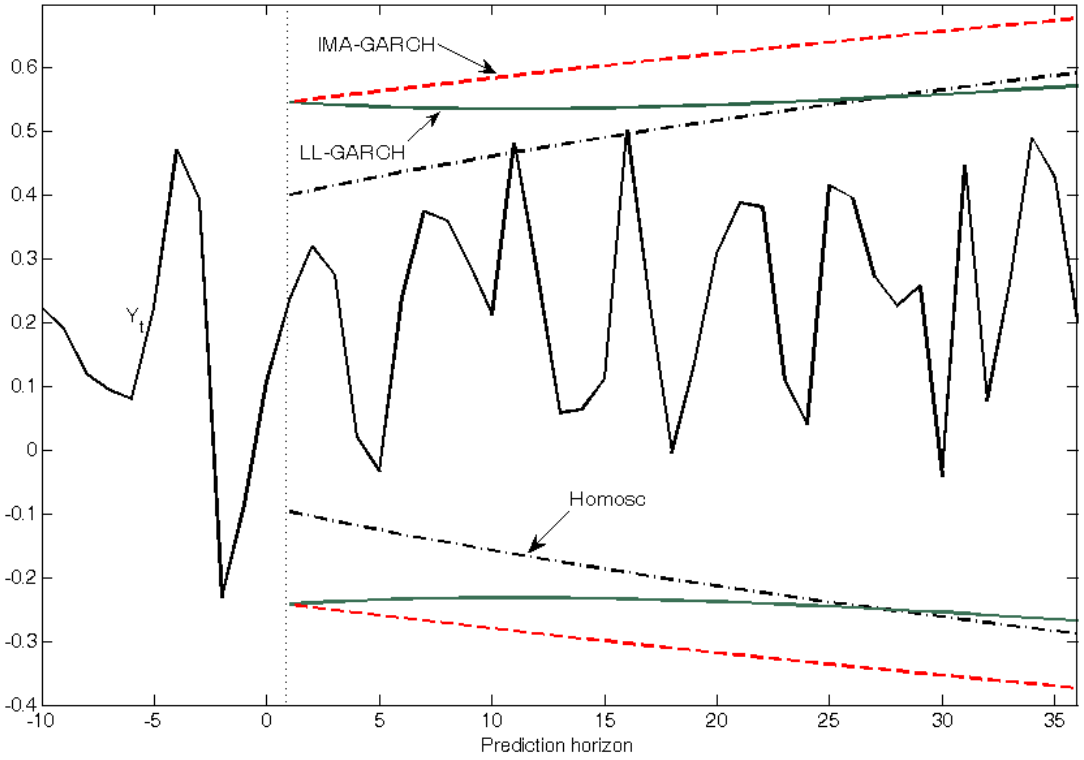

Figure 3.7: 90% Prediction intervals of the US inflation rate obtained in June 2003 for the three models considered.

Finally, we calculate the observed coverage of each prediction interval in both series. Table 3.4 reports the results for some selected horizons. For the inflation series, we choose $k = 1,\ 3,\ 6,\ 12,\ 24$ and 36. Note in this case that if the nominal coverage is 90% and the prediction horizon is $k = 1,\ 3$, or 6, the coverage of the homoscedastic intervals is smaller than the nominal, while the two heteroscedastic models have similar coverages close to 90%. However, for larger horizons, the three models generate intervals with very similar coverages, which are clearly larger than the nominal 90%. When the nominal coverage is 95%, the conclusions are similar, although the coverages of the LL-GARCH intervals are slightly closer to the nominal, even when $k = 12$ or 24. Therefore, estimating the conditionally heteroscedastic models improves the coverage, specially for short and medium horizons. As the estimated value of the signal-to-noise ratio, q_η, is very small (around 0.1), we cannot expect big differences between the LL and the IMA models (in the limiting case where $q_\eta = 0$, the local level component model collapses into a white noise process). However, estimating directly the conditionally heteroscedastic component seems to work better for medium and long horizons.

For the exchange rate series, we select $k = 1,\ 2,\ 5,\ 10,\ 20$ and 40. In this case, given that the evaluation period is mainly characterized by having negative excess volatility, i.e. the conditional variance smaller than the marginal one, the prediction intervals of the homoscedastic model overestimates the nominal coverage, for all k. With respect to the two heteroscedastic models, they produce prediction intervals with almost identical empirical coverage. This supports the result found with simulated data, where the prediction intervals of the LL-GARCH and IMA-GARCH models have a very similar performance when only the permanent component is heteroscedastic. Finally, note also that the two heteroscedastic models produce more accurate prediction intervals than the homoscedastic model for all k, because the latter is unable to capture the volatility shocks that do not vanish with k, as they come from the permanent component.

	Monthly US inflation rate					
	Homoscedastic		LL-GARCH		IMA-GARCH	
Horizon	90%	95%	90%	95%	90%	95%
$h = 1$	81.11	88.89	88.89	91.11	90.00	91.11
$h = 3$	82.95	87.50	88.64	92.05	87.50	92.05
$h = 6$	84.71	91.76	89.41	94.12	90.59	94.12
$h = 12$	91.14	92.41	91.14	94.94	92.41	96.20
$h = 24$	95.52	97.01	94.03	95.52	94.03	98.51
$h = 36$	98.18	100.00	98.18	98.18	98.18	100.00
	Daily Pound/Euro exchange rate					
	Homoscedastic		LL-GARCH		IMA-GARCH	
Horizon	90%	95%	90%	95%	90%	95%
$h = 1$	92.89	96.89	91.56	96.44	92.67	96.89
$h = 2$	92.43	95.77	90.65	95.10	90.20	94.43
$h = 5$	93.50	96.41	92.15	95.96	92.38	95.74
$h = 10$	93.88	97.96	91.38	96.60	91.16	96.15
$h = 20$	96.06	99.30	93.97	97.45	93.97	97.68
$h = 40$	96.35	100.00	93.67	96.84	93.92	97.32

Table 3.4: Empirical coverage of the Pound/Euro exchange rate and the US inflation rate, measured as the percentage of observations laying within the 90% and 95% prediction intervals. The out-of-sample period goes from January 2000 to September 2003 (450 observations) in the exchange rate and from December 2000 to May 2008 (90 observations) in the inflation series. The values of k have been selected to show relevant prediction horizons, such as a month, a quarter or a year ahead in the inflation series and a day, a week, a month or two month ahead in the exchange rate series.

3.5 Conclusions

In this chapter, we analyze the differences between prediction intervals for conditionally heteroscedastic series with stochastic trends, constructed using either unobserved component models with GARCH disturbances or the alternative ARIMA-GARCH model. In particular, we consider the local level and smooth trend models with GARCH(1,1) disturbances and their corresponding IMA-GARCH models.

We have seen that, in the local level and the smooth trend models where the transitory component, ε_t, is the only heteroscedastic component, the shocks to the variance are purely transitory. Consequently, the prediction intervals based on the homoscedastic and heteroscedastic unobserved components models stick to each other for large prediction horizons. However, depending on the sign of the excess volatility, the ARIMA-GARCH counterparts may be wider or thinner than the intervals obtained with the corresponding unobserved component model. This is due to its incapacity of distinguishing whether the heteroscedasticity affects the long or the short run components, and it may lead to significant differences between the two prediction intervals, specially for medium and long term. Therefore, the use of reduced form ARIMA models to construct prediction intervals may be inappropriate to capture the underlying uncertainty of the heteroscedastic components. These results are illustrated with simulated data and with two real time series of the Pount/Euro exchange rate and the US monthly inflation rate.

Chapter 4

Modelling inflation: conditional heteroscedasticity versus time-varying parameters

4.1 Introduction

Some recent literature on forecasting inflation has been applying unobserved component models with conditionally heteroscedastic noises to describe and predict the dynamic evolution of monthly prices; see, for example, Stock and Watson (2002, 2007), Cecchetti et al. (2007) and Broto and Ruiz (2009) for references. These models are very flexible and lead to predictions of future prices with adequate properties. However, as it was stated in previous chapters, dealing with conditionally heteroscedastic unobserved component models could be complicated due to the presence of several disturbances and the consequent difficulties involved in their estimation. Therefore, many authors represent the dynamic evolution of inflation by fitting reduced form ARIMA models with conditionally heteroscedastic errors; see, for example, Bos et al. (2007). In Chapter 2 we have seen that, although the noise of the reduced form ARIMA model is not a GARCH process, it can be well approximated by it. On the other hand, conditional heteroscedasticity has been related with ARIMA models with time-varying parameters. For example, Stock and Watson (2007), in a very interesting and influential paper, propose to represent the inflation by the local level model with non-stationary stochastic volatility disturbances. They argue that the presence of heteroscedasticity in the components, generates a reduced form MA(1) model with a time-varying MA parameter. When the models for the conditional variances of the unobserved components are stationary, we show in Chapter 2 that the parameters of the conditional mean of the reduced form ARIMA model are constant over time. This apparent contradiction may be explained by the fact that Stock and Watson (2007) are assuming a non-stationary heteroscedastic model for the unobserved noises, while our results are given under the assumption of stationarity. In general, if the marginal variances are constant, the presence of conditional heteroscedasticity in the components does not affect the signal-to-noise ratio. Consequently, it does not affect the parameters governing the conditional mean of the process. Therefore, the

point predictions obtained by fitting the unobserved component and the reduced form ARIMA models are similar. However, structural breaks in the marginal variances of the components can generate changes in the signal to noise ratios and, consequently, the reduced form parameters will also change.

In this chapter, we focus on the local level model and analyze the evolution of the reduced form MA parameter when the marginal variances of the local level noises have structural breaks. We consider two different assumptions about the conditional second order moments of the noises: i) they are homoscedastic, ii) they are conditionally heteroscedastic but stationary. We show that the evolution of the MA parameter, θ, detected by Stock and Watson (2007) can be only attributed to breaks and not to the stationary evolution of the conditional variances. However, given that the random walk model assumed for the log-volatilities of the components is well designed to deal with structural breaks, they are able to explain changes in the MA parameters with the conditionally heteroscedastic local level model.

We also perform an empirical analysis of the seasonally adjusted US inflation rate, measured as the personal consumption expenditure deflator (PCE). We fit different unobserved component and ARIMA models to the series. Then, we obtain out-of-sample point and interval predictions and evaluate the relative forecasting performance of both models. The point forecast evaluation has already been performed by many authors; see, for example, Nelson and Schwert (1977), Atkeson and Ohanian (2001), Fisher et al. (2002), Orphanides and van Norden (2005) and Stock and Watson (2007). In this chapter, we also measure the forecasting performance by assessing the accuracy of prediction intervals, which may be as important as the point predictions, specially for policy makers. We show that the heteroscedastic models provide point and interval forecasts that are more accurate than their homoscedastic counterparts. Additionally, the local level model with the transitory component being GARCH(1,1) and the noise of the permanent component having a break in its marginal variance, produces the most accurate point forecasts for almost all horizons.

The rest of the chapter is structured as follows. In Section 4.2, we briefly describe the local level model with heteroscedastic disturbances and then we study with simulated data the properties of the rolling window estimate of the IMA parameter coming from series with highly persistent noises in either or both unobserved components. We focus on whether the time-varying IMA model is able to capture the dynamics of the evolving volatility given by the conditionally heteroscedastic local level model. To this end, we carry out several Monte Carlo exercises. In Section 4.3, we perform a forecasting exercise with a US inflation series to study whether fitting unobserved component models will provide more accurate results in terms of point and interval forecasts, when compared with the rolling estimation or some other reduced form IMA models. Finally, Section 4.4 concludes.

4.2 The local level model and the rolling window IMA approach

One useful model to represent the dynamics of seasonally adjusted inflation is the local level model. Nelson and Schwert (1977) is an early reference that shows the good forecasting per-

formance of the IMA model (the reduced form of the local level model) fitted to monthly US inflation observed from 1953 to 1971. More recently, Broto and Ruiz (2009), Cogley and Sargent (2007), Müller and Watson (2006) and Stock and Watson (2002, 2007) show the usefulness of the local level model for forecasting US inflation. Recall that the local level model assumes that the series of interest, y_t, is composed by an underlying stochastic level that contains a unit root, μ_t, and a transitory component, ε_t, as in (1.2). The reduced form of this model is obtained by looking for the ARMA model that has the same autocorrelation function of Δy_t. If σ_ε^2 and σ_η^2 are both finite, then the acf of Δy_t is given by (2.5). Note that q_η (from now on, just q), obtained as the ratio of the marginal variances of the long run and transitory components, plays a fundamental role in determining the size of the first order autocorrelation of Δy_t. By looking at the acf of the local level model, it is straightforward to show that the reduced form model of y_t with a unique disturbance is an ARIMA(0,1,1) whose parameters are determined by q and σ_ε^2 in the following way

$$\theta = \frac{(q^2+4q)^{1/2}-2-q}{2}, \tag{4.1a}$$

$$\sigma_a^2 = -\frac{\sigma_\varepsilon^2}{\theta}. \tag{4.1b}$$

Note that if the marginal variances are finite and constant over time in (4.1), then the MA parameters will also be constant. However, we cannot ensure that either θ or σ_a^2 remain constant whenever some of the noises in the local level model are nonstationary in second moments. In this section we study this issue with the aid of simulated data.

4.2.1 Constant marginal variances

In Chapter 2 we have already illustrated with simulated time series that when the local level noises have finite and constant marginal variances then the value of θ is constant, regardless of whether these noises are conditionally heteroscedastic. As a corollary of this result, we can conclude that when the parameter θ evolves over time, this evolution cannot be attributed to the presence of stationary conditional heteroscedasticity in the components but to structural breaks in the marginal variances or to other causes generating non-constant marginal variances.

Obviously, even when the parameter θ is truly constant over time, if we estimate this parameter with a rolling window, we will obtain estimates that evolve over time around θ. To illustrate the kind of evolution that could be observed, we generate four series of size $T = 500$ with the homoscedastic local level model and find the rolling window estimates of the MA parameter, $\hat{\theta}_{rol}$, using a window of 100 observations. We use this setup to resemble the case of a monthly series observed during approximately 40 years, and using a rolling window of around eight years, commonly used for fitting inflation series. Figure 4.1 shows these estimates, together with the implied θ coming from the estimation of q. The bands correspond to ± 2 asymptotic standard deviations of $\hat{\theta}_{rol}$. We choose $q = 0.5$ ($\theta = -0.5$) for this exercise. Note that the resulting $\hat{\theta}_{rol}$ have a very persistent dynamic behavior, resembling a nonstationary process. In fact, the larger the window size, the higher the autocorrelations between consecutive values of $\hat{\theta}_{rol}$ and therefore the more persistent is the whole process. This happens because a shock at a given time point on

the series will persist for at least 100 values of the $\hat{\theta}_{rol}$ until the rolling window does not include this observation.

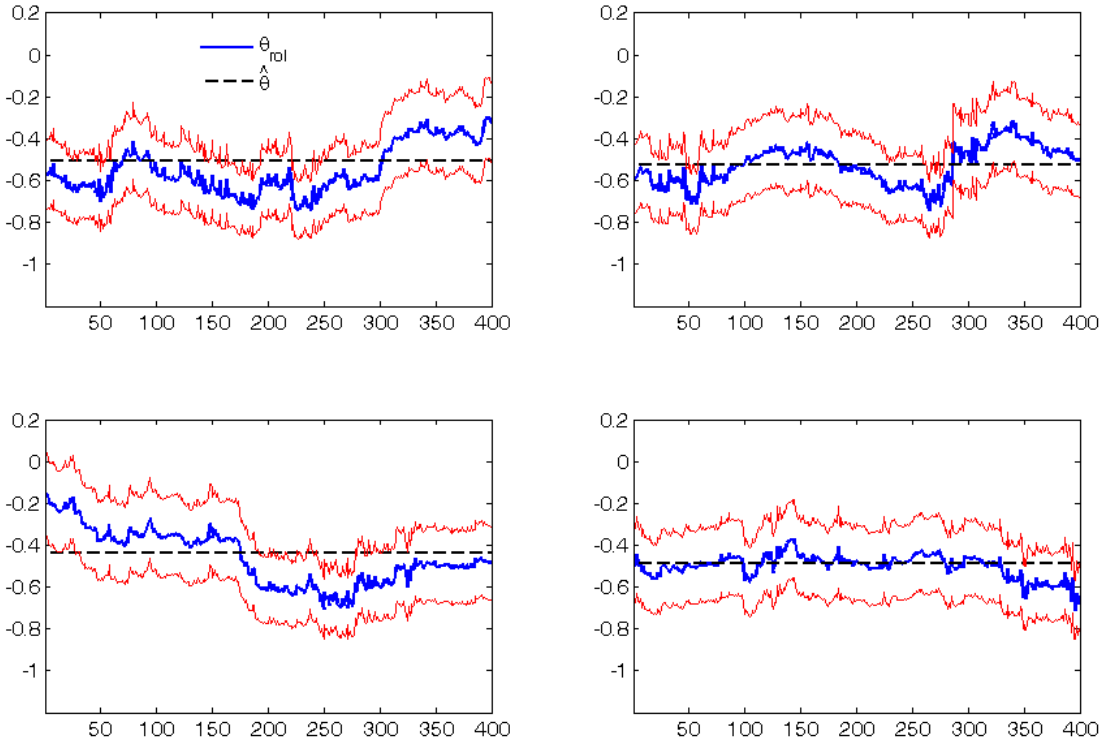

Figure 4.1: Rolling window MA estimate, $\hat{\theta}_{rol}$, and the implied θ of series simulated from the homoscedastic local level model with $q = 0.5$ ($\theta = -0.5$). The bands correspond to ± 2 standard deviations of $\hat{\theta}_{rol}$.

The above results are obtained regardless of the noises being homoscedastic or heteroscedastic, as far as the series are stationary with constant marginal variances. If, in particular, we consider that either or both noises are stationary GARCH or ARSV processes, then, given that the marginal variances are constant, we would obtain the same type of behavior as in the homoscedastic case. For example, Figure 4.2 plots the rolling window estimates and the implied θ that comes from the estimation of q, using either the heteroscedastic version of the Kalman Filter, $\hat{\theta}_{(het)}$, proposed by Harvey et al. (1992), or the homoscedastic one, $\hat{\theta}_{(hom)}$. The results correspond to the local level series with GARCH noises in the transitory component, ε_t, and whose parameters are $\alpha_0 = 0.05$, $\alpha_1 = 0.15$, $\alpha_2 = 0.8$, $\gamma_0 = 0.5$, i.e. a marginal $q = 0.5$ and thus $\theta = -0.5$. From the plots we observe that the behavior of the rolling window estimates track always the marginal θ, as in the homoscedastic cases. Note also that the estimates, $\hat{\theta}_{(het)}$ and $\hat{\theta}_{(hom)}$, are pretty similar to each other and close to the real value, -0.5, thus indicating that both filters provide accurate estimators of the marginal variances, when they exist and are constant over time.

4.2.2 Nonstationary processes and structural breaks in the local level noises

The above simulations have been carried out assuming finite and constant marginal variances. Therefore, the evolution of $\hat{\theta}_{rol}$ observed in Figures 4.1 and 4.2 can be attributed to sampling

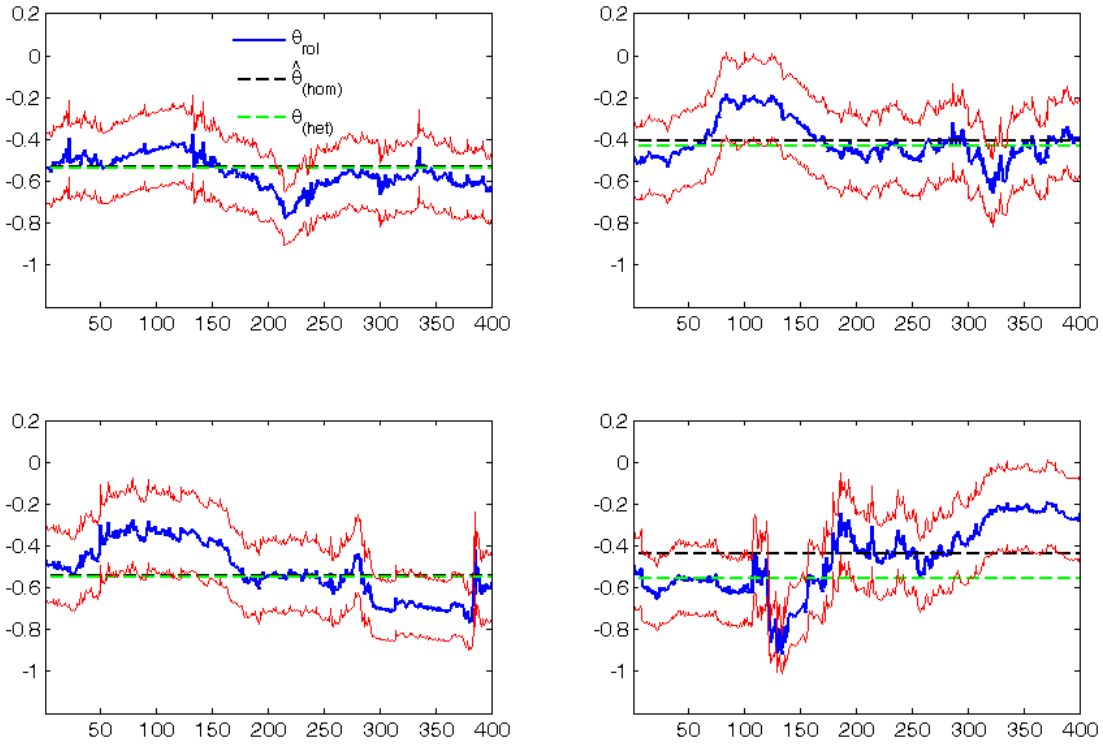

Figure 4.2: Rolling window estimates of the MA parameter, $\hat{\theta}_{rol}$, and the two estimates of the implied θ, according to whether the homoscedastic, $\theta_{(hom)}$, or the heteroscedastic, $\theta_{(het)}$, filter is used. The series are generated from the local level model with GARCH noises where $\alpha_0 = 0.05$, $\alpha_1 = 0.15$, $\alpha_2 = 0.8$, $\gamma_0 = 0.5$ and $\gamma_1 = \gamma_2 = 0$. The bands correspond to ± 2 standard deviations of $\hat{\theta}_{rol}$.

variability of the QML estimator of θ. If we now consider nonstationary heteroscedastic models for the noises, ε_t and η_t, then, θ is not defined. However, it may be interesting to estimate it anyway in order to give an insight about the shape of the rolling window estimates. Thus, we generate series with IGARCH processes on ε_t and η_t, and obtain the $\hat{\theta}_{rol}$. In the local level model, the IGARCH process for each noise is obtained when $\alpha_0 = \gamma_0 = 0$ and $\alpha_1 + \alpha_2 = \gamma_1 + \gamma_2 = 1$ in (2.8). Figure 4.3 plots the rolling windows estimates of θ for four particular time series. In the examples it is evident that the dynamic behavior of $\hat{\theta}_{rol}$ is very persistent and, more importantly, it does not in general evolve around a specific value. However, this dynamics is not related to the presence of breaks in the data, and it is simply the consequence of undefined marginal variances. Furthermore, note that by just looking at the evolution of $\hat{\theta}_{rol}$, it is very difficult to decide whether this evolution can be attributed to sample variability of the estimator of θ or it is the consequence of a non-stationary evolution of the conditional variance.

On the other hand, given that θ is a function of the marginal variances, if any of them has a structural break, then θ captures it and shows also a break. In order to illustrate this effect, we now generate series from the homoscedastic and heteroscedastic local level model with breaks in the marginal variances. The size and time of these breaks are chosen to replicate those usually found in the US inflation; see, for example, Sensier and van Dijk (2004). The sample size is set equal to 557 and the unique break is introduced in observation 299 in the marginal variance of both noises. Moreover, the size of the shock has been set in a way that the signal-to-noise ratio decreases from $q = 2.4$ to $q = 0.25$ after the break, thus replicating the shock estimated

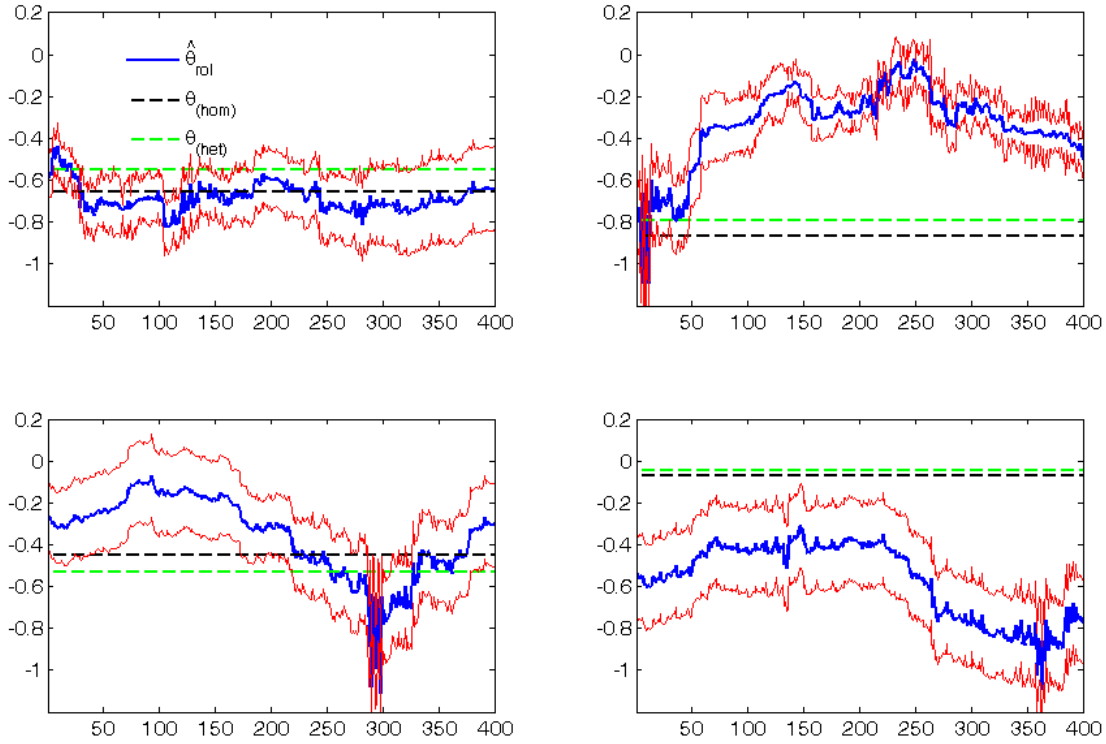

Figure 4.3: Rolling window estimates of the MA parameter, $\hat{\theta}_{rol}$, and the two estimates of the implied θ, according to whether the homoscedastic, $\theta_{(hom)}$, or the heteroscedastic, $\theta_{(het)}$, filter is used. The series are simulated from the local level model with IGARCH disturbances, with $\alpha_1 = \gamma_1 = 0.1$. The bands correspond to ± 2 standard deviations of $\hat{\theta}_{rol}$.

for the US quarterly inflation series in Stock and Watson (2007). Consequently, the random walk process of the underlying level become relatively less important in the series, a common feature shared by many US macroeconomic variables from the early eighties (often referred as the "Great Moderation"). Figure 4.4 plots the evolution of $\hat{\theta}_{rol}$ for four series simulated with homoscedastic noises and breaks as just described. Looking at the plots, there seems to be a common pattern in the evolution of the time-varying estimate of the MA coefficient. Although the rolling window estimate capture the effect of the break, it adapts very smoothly to the new level. This happens because the window size of 100 observations makes the transition too smooth. Therefore, not until 100 observations after the break the rolling window estimate of θ will be obtained entirely with observations coming from the new data generating process. In a monthly series, this means that the new level for the MA parameter will be achieved after 8 years, something unrealistic for macroeconomic data.

We also generate series in which the long run noise, η_t, is homoscedastic and the transitory noise, ε_t, is heteroscedastic. In this case, the break is introduced only in the marginal variance of the homoscedastic permanent component. This model could be useful to describe the time-varying volatility of the US inflation, as well as the effect of the "Great Moderation". Plots of the implied and rolling MA parameter are given in Figure 4.5. In general, the conclusion from the plots are the same as in Figure 4.4, in that the rolling window estimate adapts very smoothly to the new level after the breaks. Again, as in the other cases commented before, it is

very difficult to identify by just looking at the evolution of $\hat{\theta}_{rol}$, the date and size of the break.

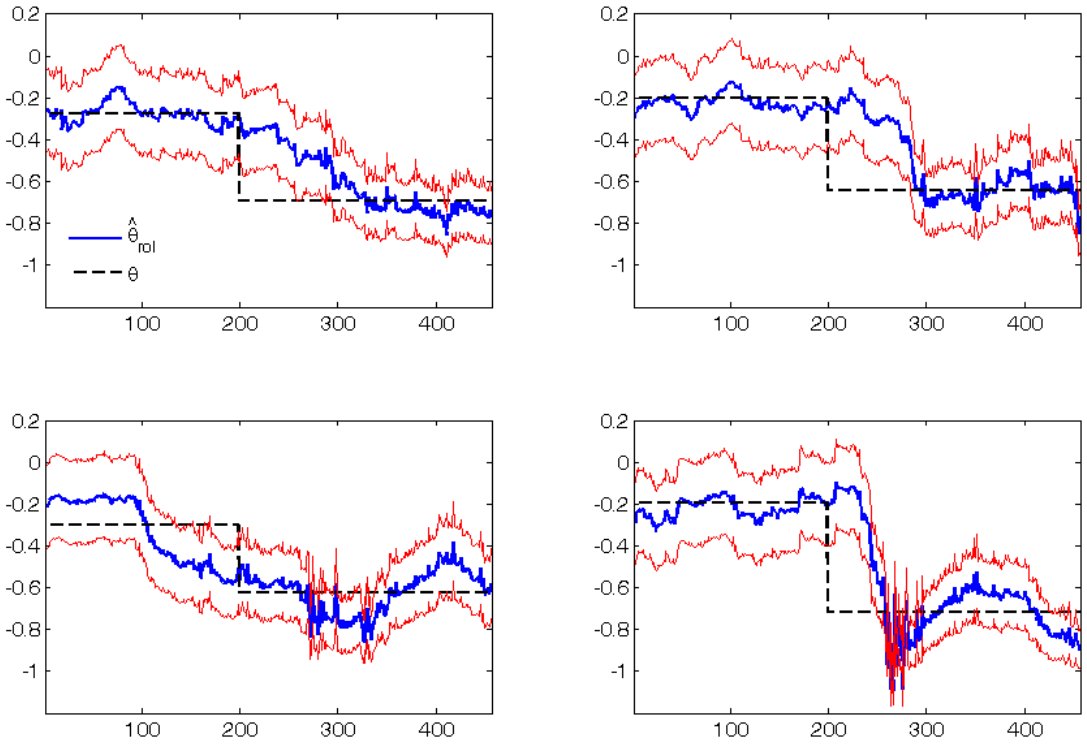

Figure 4.4: Rolling window MA estimate, $\hat{\theta}_{rol}$, and the implied θ of series simulated from the local level model with a break in its homoscedastic disturbances. The break is in observation 299, and the size in both variances is such that before this observation, $q = 2.4$ and after $q = 0.25$. The bands correspond to ± 2 standard deviations of $\hat{\theta}_{rol}$.

Alternatively, the effect of a break in either or both variances can be handled by imposing integrated processes for the variances of the local level noises and then let the conditional variances to adapt accordingly, as in Stock and Watson (2007). This model should be flexible enough to incorporate the discrete jump more quickly. As an illustration, in Figure 4.6 we show the estimated $\sigma^2_{\eta,t}$ for the local level model with an IGARCH process for η_t to capture the breaks in the variance of this noise present in the series given in Figure 4.5. Indeed, we can see from the plots that the estimated conditional variance of η_t captures the break and accommodates to the new levels very quickly, in opposition to the evolution of the rolling windows estimate of θ. In the plots of Figure 4.6 we also include an estimate of $\sigma^2_{\eta,t}$ resulting from a random walk stochastic volatility (RW-SV) process on η_t. Note that this model yields estimates of $\sigma^2_{\eta,t}$ that are very similar to those of the IGARCH model.

In any case, comparing Figures 4.1 to 4.5, we can conclude that it is very difficult to assess if the evolution of the $\hat{\theta}_{rol}$ comes only from sample variability or from changes in the marginal variances of the local level noises.

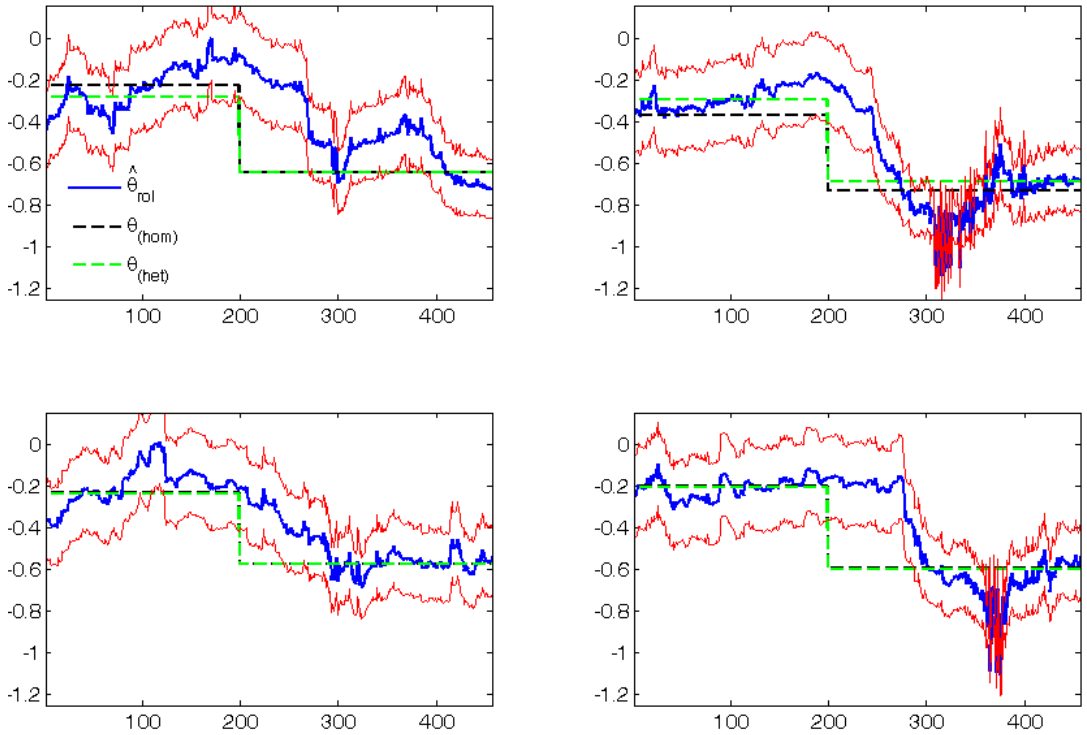

Figure 4.5: Rolling window MA estimate, $\hat{\theta}_{rol}$, and the implied θ of series simulated from the local level model with a break in the variance of the homoscedastic η_t and with a GARCH process in ε_t, in order to mimic the dynamics of the US monthly inflation.

4.2.3 Conditional volatility in the rolling windows IMA model

In the last subsections, we have discussed the relationship between the local level model with conditionally heteroscedastic noises and/or breaks in the underlying marginal variances, and the rolling window estimate of θ in the reduced form IMA model. In this subsection, we study the properties of the residuals resulting from the time-varying IMA model, a_t. In particular, we analyze whether the the rolling windows IMA model is able to capture the conditional heteroscedasticity coming from the local level noises.

In order to study the properties of the residuals of the rolling window IMA model, Table 4.1 reports results of some diagnostic tests for 100 simulated series generated by the local level models considered above. For each model, it reports the percentage of cases the 10th-order Ljung-Box statistic (Q(10)), the Jarque-Bera (J-B) normality test, the 10th-order Rodriguez and Ruiz (2005) test statistic ($Q_1(10)$) to the squared residuals, and the LM test for ARCH effects with j lags (ARCH(j)) reject the corresponding null hypothesis at 5% significance level. The estimates of a_t for the rolling window estimation approach are recovered in a recursive way from

$$\hat{a}_t \quad = \quad \Delta y_t - \hat{\theta}_{rol,t}\, \hat{a}_{t-1}, \tag{4.2}$$

where $\hat{\theta}_{rol,t}$ is the QML estimate of the MA parameter for the sample $t-99, ..., t$. First note that the Q(10) test statistic is around the nominal 5% only in the cases where the noises are homoscedastic. However, if some of the noises are heteroscedastic, this test over rejects the null

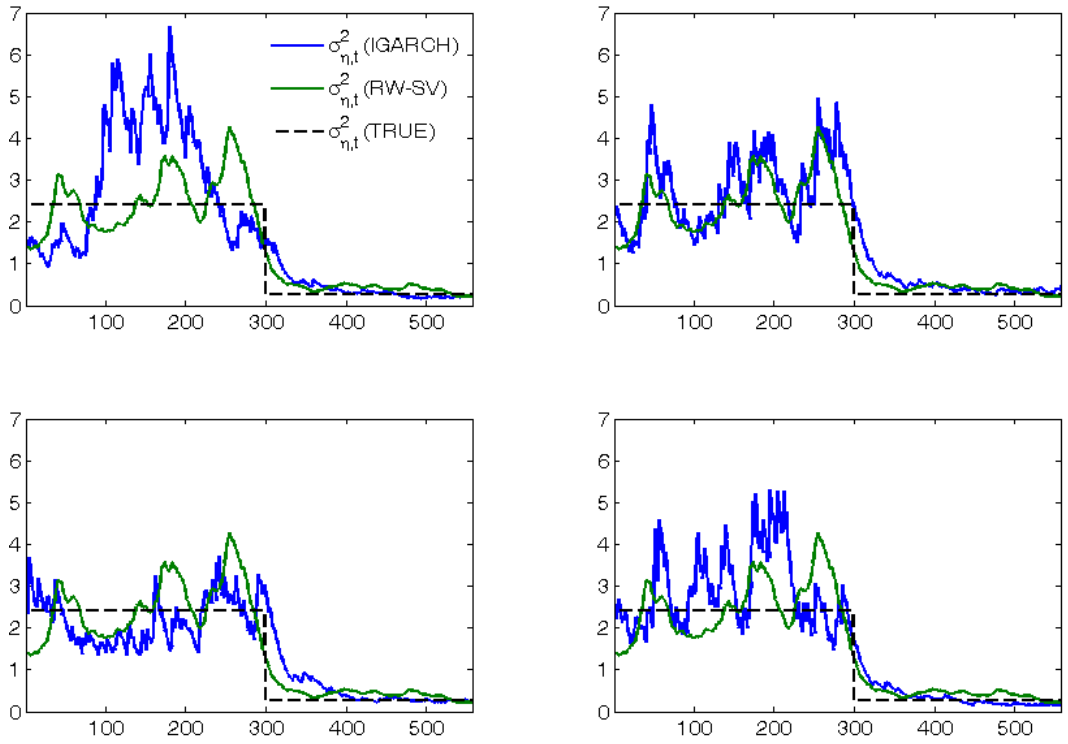

Figure 4.6: Estimated time-varying volatility of η_t, $\sigma^2_{\eta,t}$, resulting from two different non-stationary heteroscedastic models, the IGARCH and the random walk stochastic volatility models, and fitted to capture the break in the marginal variance of η_t. The size and location of the break are set to replicate the dynamics of the US inflation.

of zero autocorrelation, i.e. the size of the test is too large. On the other hand, the $Q_1(10)$ and ARCH(j) tests show that the residuals obtained from (4.2) have significant autocorrelations of squares when some of the local level noises are heteroscedastic. This means that allowing the estimator of θ to change over time by using a rolling window approach does not capture the heteroscedastic nature of the local level disturbances. On the other hand, the presence of breaks in the variances of the local level noises does not lead to conditionally heteroscedastic innovations in the time-varying IMA models. Finally, note that when both noises are IGARCH, the marginal variances and covariances of a_t are not defined and, consequently, the tests are no longer valid.

4.3 Empirical evidence with the US inflation rate

In this section, we fit several models to the US seasonally adjusted monthly inflation rate. These models are chosen to capture different features of the series that may affect the forecasting performance. Nextly, we perform an out-of-sample forecast evaluation of the point and interval predictions generated by these models.

DGP on the local level	Q(10)	J-B	$Q_1(10)$	ARCH(1)	ARCH(4)
$\varepsilon_t \sim$ GARCH $\eta_t \sim$ Homosc	12.7%	40.2%	67.6%	32.5%	53.2%
$\varepsilon_t \sim$ Homosc $\eta_t \sim$ Homosc	3.4%	5.1%	5.8%	5.3%	5.8%
$\varepsilon_t \sim$ IGARCH $\eta_t \sim$ IGARCH	89.6%	84.8%	99.4%	94.7%	98%
$\varepsilon_t \sim$ Homosc w/breaks $\eta_t \sim$ Homosc w/breaks	5.1%	6%	5.6%	3.6%	5.4%
$\varepsilon_t \sim$ GARCH $\eta_t \sim$ Homosc w/breaks	10.8%	36.7%	70.2%	26.8%	48.5%

Table 4.1: Diagnostic tests for the residuals of the rolling window IMA model fitted to series coming from the local level model. The figures reported are the percentage of cases the 10th-order Ljung-Box statistic (Q(10)), the Jarque-Bera (J-B) normality test, the 10th-order Rodriguez and Ruiz (2005) test statistic ($Q_1(10)$) to the squared residuals, and the LM test for ARCH effects with j lags (ARCH(j)) reject the corresponding null hypothesis at 5% significance level.

4.3.1 The Data

We use the Personal Consumption Expenditure (PCE) deflator as a measure of inflation. Calling P_t the price index at time t, we define the inflation rate as $\Pi_t = 100 \times \log(P_t/P_{t-1})$. The observation period spans from 02/1959 to 05/2008. This leads to a sample of $T = 592$. The first 332 observations are used for the estimation, while the last 60 observations are left for the out-of-sample forecast evaluation. The series has been downloaded from the EconWin database, and it is already seasonally adjusted. Figure 4.7 plots the inflation series. As suggested by many authors, the "Great Moderation effect" from the beginning of the eighties onwards seems to make the series less erratic.

With respect to the descriptive statistics, Table 4.2 reports some sample moments of the first difference of the inflation series, $\Delta\Pi_t$. In order to have more information about the behavior of the series prior and after the "Great Moderation", we also split the sample and compute the descriptive statistics, separately for the periods spanning from 02/1959 to 12/1983 and from 01/1984 to 05/2008. The date for the breakpoint is chosen according to what is found in the literature as the beginning of the Great Moderation; see, for example, Sensier and van Dijk (2004) and Stock and Watson (2007) and the references therein. First, note that the marginal variances of the series does not vary significantly in the two periods. On the other hand, when looking at the sample autocorrelations on the first column of Figure 4.8, we see that $r(1)$ is significant and negative, with slightly differences between the two periods and the estimated for the whole sample. Note also that $r(2)$ is also significant in the second period. The second column of this figure shows the differences between the sample autocorrelations of squares and the squared sample autocorrelations, $\bar{r}_2(\tau) = r_2(\tau) - [r(\tau)]^2$. If we approximate the standard

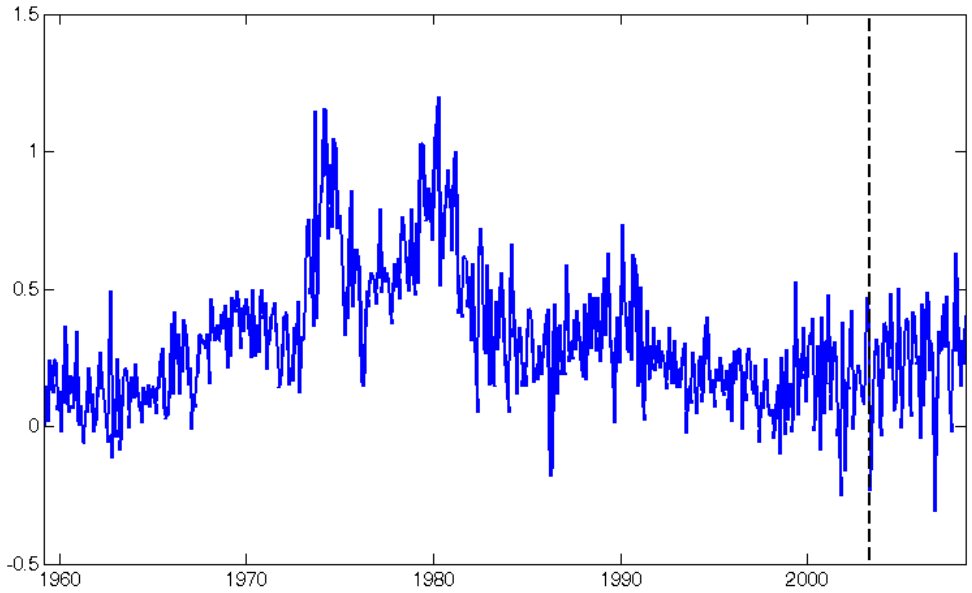

Figure 4.7: Seasonally adjusted monthly inflation series, defined as $100 \times \log(P_t/P_{t-1})$, where P_t is the price index coming from the Personal Consumption Expenditure (PCE) at month t. The sample period goes from 02/1959 to 05/2008. The vertical dashed line corresponds to 06/2003, the starting point of the out-of-sample forecast evaluation period.

deviation of $\bar{r}_2(\tau)$ to be $1/\sqrt{T}$, we see that there exist some significant values of $\bar{r}_2(\tau)$. This fact, together with the presence of some positive excess kurtosis, suggest that the series may present a conditional heteroscedastic structure that should be considered, specially in the first period.

	Mean	Var	Skew	Exc Kurt
First Period: 02/1959-12/1983	0.001	0.030	-0.228	2.832
Second Period: 01/1984-05/2008	0.000	0.036	-0.033	0.470
Whole Sample: 02/1959-05/2008	0.001	0.033	-0.120	1.492

Table 4.2: Descriptive statistics of the first difference of the inflation rate. The analysis is performed for the whole period, 02/1959-05/2008, and also for the part of the sample prior to and after 01/1984.

4.3.2 Model selection and estimation

The correlations reported in Figure 4.8 suggest that an MA(1) with negative parameter or alternatively, the random walk plus noise model can be appropriate both when we consider the whole sample period or when we look at each of the subperiods separately. Consequently, we start by fitting these two models with homoscedastic noises. Table 4.3 reports the QML estimates of the parameters θ and q. The differences in the values of $\hat{q}$ prior to and after 01/1984 suggest that the "Great Moderation" may involve a structural break in the variances of the local level noises and should be taken into account when selecting the models to fit these series. Moreover, the results show that only the noise of the permanent component has a significant break in its marginal variance. Then, it would be worthy to add a model that includes a discrete jump that

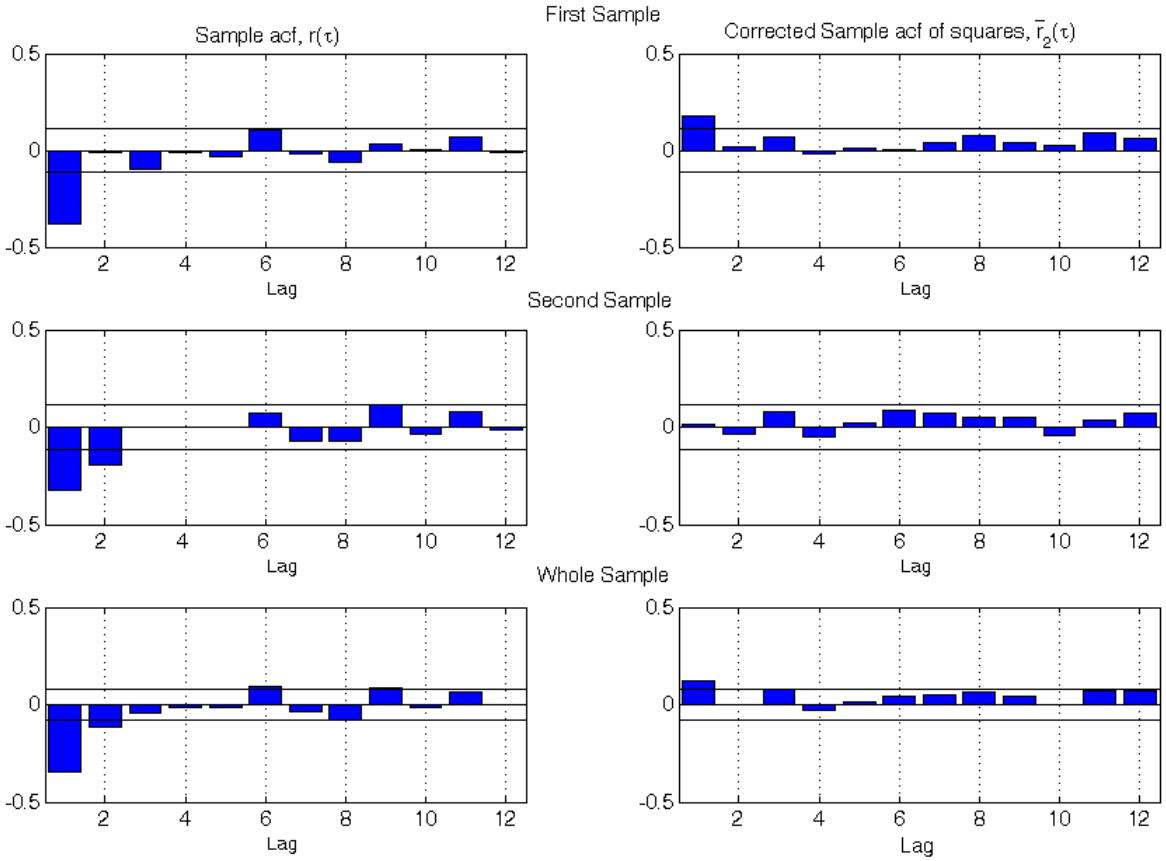

Figure 4.8: Sample autocorrelations, $r(\tau)$, and the differences between the sample autocorrelations of squares and the squared sample autocorrelations, $\bar{r}_2(\tau) = r_2(\tau) - [r(\tau)]^2$, for the first difference of the inflation series. The analysis is performed for the whole period, 02/1959-05/2008, and also for the sample periods prior to and after 01/1984.

captures this break. On the other hand, we also observe a significant change in $\hat{\theta}$ in the IMA model, but with a marginal variance of a_t constant between the two sample periods. As done in Stock and Watson (2007), we use a 10-year rolling-windows approach to capture the changes on θ.

In order to capture the heteroscedastic structure that may be present in the inflation process, we also add a model in which the local level noises are GARCH. Following the same argument as in Chapter 3, we decide to assume a GARCH process only in the transitory component. Additionally, if we want also to capture possible breaks in the marginal variances of the noises, we need to include a discrete jump in the noise of the permanent component as in the homoscedastic case. Alternatively, we also fit a GARCH process on the residuals $\hat{a}_t$ for the IMA models with θ constant and with the 10-year rolling window estimate, $\hat{\theta}_{rol}$. Finally, we also fit two other models to the series that include non-stationary processes in the form of either IGARCH or RW-SV model, as an alternative way of capturing possible breaks in the local level variances. Analogously, we also fit an IMA model with IGARCH noises. Thus, summarizing, we fit six different local level models, and other six IMA models, in order to capture the features of the series that lead to improve the forecasting performance. Namely, the twelve models, with their labels in parenthesis, are given by

Local level models				IMA models			
$\hat{q}_{(all)}$	$\hat{q}_{(1st)}$	$\hat{q}_{(2nd)}$	$\hat{q}_{(2nd)} - \hat{q}_{(1st)}$	$\hat{\theta}_{(all)}$	$\hat{\theta}_{(1st)}$	$\hat{\theta}_{(2nd)}$	$\hat{\theta}_{(2nd)} - \hat{\theta}_{(1st)}$
0.081	0.167	0.009	-0.158**	-0.781	-0.660	-0.899	-0.240**
$\hat{\sigma}^2_{\varepsilon,(all)}$	$\hat{\sigma}^2_{\varepsilon,(1st)}$	$\hat{\sigma}^2_{\varepsilon,(2nd)}$	$\hat{\sigma}^2_{\varepsilon,(2nd)} - \hat{\sigma}^2_{\varepsilon,(1st)}$	$\hat{\sigma}^2_{a,(all)}$	$\hat{\sigma}^2_{a,(1st)}$	$\hat{\sigma}^2_{a,(2nd)}$	$\hat{\sigma}^2_{a,(2nd)} - \hat{\sigma}^2_{a,(1st)}$
17.2	15.3	19.5	4.2	24.3	23.0	24.1	1.1
$\hat{\sigma}^2_{\eta,(all)}$	$\hat{\sigma}^2_{\eta,(1st)}$	$\hat{\sigma}^2_{\eta,(2nd)}$	$\hat{\sigma}^2_{\eta,(2nd)} - \hat{\sigma}^2_{\eta,(1st)}$				
1.4	2.5	0.2	$-2.3**$				

Table 4.3: Quasi maximum likelihood (QML) estimates of σ^2_ε, σ^2_η and q of the local level model and θ and σ^2_a of the IMA model, both assuming homoscedastic noises. The subindexes (all), $(1st)$ and $(2nd)$ correspond to the estimates for the whole, first and second sample periods, respectively, i.e. prior to and after 01/1984. The variances are multiplied by 10^3. *(**) means significant at 5% (1%).

1. Local level model with homoscedastic noises and without breaks (*Homoscedastic w/o breaks*).
2. Local level model with homoscedastic noises and with step breaks in the variances of both, ε_t and η_t in 01/1984 (*Homoscedastic w/ breaks*).
3. Local level model with GARCH(1,1) process in ε_t and homoscedasticity in η_t (*GARCH in ε_t w/o breaks*).
4. Local level model with GARCH(1,1) process in ε_t, and homoscedasticity and a break in 01/1984 in η_t (*GARCH in ε_t w/ breaks*).
5. Local level model with IGARCH(1,1) processes in both noises (*IGARCH in both noises*).
6. Local level model with a stationary ARSV process in ε_t and a RW-SV in η_t (*ARSV in ε_t and RW-SV in η_t*)
7. IMA(1,1) model with homoscedastic a_t (*Homoscedastic*)..
8. IMA(1,1) model with a GARCH(1,1) process in a_t (*GARCH in a_t*).
9. IMA(1,1) model with a IGARCH(1,1) process in a_t (*IGARCH in a_t*).
10. 10-year rolling window IMA(1,1) model with homoscedastic a_t (*S & W approach*).
11. 10-year rolling window IMA(1,1) model with a GARCH(1,1) process in a_t (*GARCH in a_t*).
12. 10-year rolling window IMA(1,1) model with a IGARCH(1,1) process in a_t (*IGARCH in a_t*).

Given that we leave the last 60 observation for out-of-sample evaluation, we first estimate the models from 02/1959 to 05/2003 (532 observations) and then re-estimate them moving the sample one month at a time. For the 10-years rolling window IMA models, the first sample

is from 06/1993 to 05/2003. At each step, the point and interval forecasts at three different horizons, $h = 1$, 6 and 12 months, are recorded. Finally, these values are used to evaluate the forecasting performance of each model. We use quasi-maximum likelihood (QML) estimation methods in all models but the local level with ARSV noises. In this case, we use MCMC methods to obtain posterior distributions of the parameters and the unobserved components[1]. Table 4.4 reports the p-value of the Ljung-Box statistic, $Q(10)$, and the Rodriguez and Ruiz (2005) statistic, $Q_1(10)$, on the level and squared standardized residuals, respectively. In general, the fit of the conditional mean seems to produce "clean" residuals only in the 10-years rolling window IMA models. On the other hand, when analyzing the squared residuals, we see that only the heteroscedastic models eliminate the correlation structure.

Local Level	$Q(10)$	$Q_1(10)$
Homoscedastic w/o breaks	0.006	0.000
Homoscedastic w/ breaks	0.008	0.000
GARCH in ε_t w/o breaks	0.006	0.175
GARCH in ε_t w/ breaks	0.009	0.270
IGARCH in both noises	0.005	0.393
ARSV in ε_t and RW-SV in η_t	0.002	0.052
IMA(1,1)		
Homoscedastic	0.006	0.000
GARCH in a_t	0.003	0.245
IGARCH in a_t	0.002	0.571
Rolling Window TV-IMA(1,1)		
S & W approach	0.280	0.420
GARCH in a_t	0.408	0.981
IGARCH in a_t	0.392	0.951

Table 4.4: Diagnosis for the standardized residuals of the models selected to fit the inflation series. The sample covers the fist $T - R = 532$ observations, from 02/1959 to 05/2003. The figures are the p-values of the 10th-order Ljung-Box statistic, $Q(10)$, applied to the level residuals and the 10th-order Rodriguez and Ruiz (2005) statistic, $Q_1(10)$, applied to the squares residuals.

4.3.3 Forecasting performance of the competing models

After re-estimating the models for the whole evaluation period, we obtain the point and intervals forecasts for each horizon h. Table 4.5 summarizes the information with respect to point forecasts for $h = 1$, 6 and 12. The values correspond to the root mean squared forecast error (RMSFE) and mean absolute deviation (MAD) for each model and horizon, relative to the 10-year rolling window IMA(1,1) model with homoscedastic a_t, i.e. the Stock and Watson (2007) approach,

[1]To obtain the posterior distributions, we use WinBugs, a package developed for doing Bayesian inference, specially through MCMC methods; see Meyer and Yu (2000). A free version of this software is available at *www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml*

which we use as the benchmark model. Thus, models with values of the RMSFE or MAD less than one means that they produce more accurate point forecasts than the homoscedastic 10-year rolling window model. We check the significance of these differences with the test proposed by Diebold and Mariano (1995). This test performs a pairwise comparison of the accuracy of two models, with respect to a given loss function. We use quadratic differences of the forecast errors in the columns corresponding to the RMSFE, and absolute differences of the forecast errors in the columns corresponding to the MAD. Thus, for instance, a value with *(**) in the RMSFE columns means that the point forecast of the corresponding model is significantly more accurate than the benchmark model at 5%(1%). Note that for the three horizons, there is at least one model that significantly improves the accuracy of the benchmark model. Moreover, note that the best point-forecasts (measured as the lowest RMSFE and MAD) in almost all cases is the heteroscedastic local level model with the deterministic break in the variance of the permanent component. Observe also that the models with non-stationary specifications for the conditional variances in the local level models, do not yield more accurate point forecasts than the other local level models. On the other hand, the IMA models fitted to the whole sample and with θ fixed yield in general worse point forecasts than the local level models, although in this case, assuming an integrated GARCH process in the residuals seems to improve forecast accuracy. With respect to the 10-years rolling window models, their accuracy are about 10 or 15% worse than the local level models with the best performance. Finally, among the local level models, note that imposing a deterministic break to account for the change in the marginal variances improves forecasting accuracy.

Table 4.6 reports the empirical coverage, calculated as the percentage of observations laying inside the 90% and 95% intervals for each model and horizon. First, note that in general, the coverage is underestimated. With respect to the different models, it seems that the integrated versions in both, the local level and IMA models, produce prediction intervals with coverages closer to the nominal. However, when disregarding these models, again the heteroscedastic models yield a coverage closer to the nominal than the homoscedastic ones. This implies that the models with homoscedastic noises produce in general worse coverages than those with conditionally heteroscedastic noises.

Summarizing the results regarding forecasting performance in the inflation series for the period considered, we conclude that the conditionally heteroscedastic models produce more accurate point and interval forecast than those with homoscedastic noises. Moreover, the results on the point forecast evaluation suggest that the conditionally heteroscedastic local level model is in general better than the IMA counterparts, and that the inclusion of the deterministic break in the marginal variance of η_t significantly improves forecast accuracy, for almost all horizons. However, in what concerns to interval forecasting evaluation, results are no so clear, and there is not a dominant model for all series. In any case, the models with conditionally heteroscedastic noises seem to generate prediction intervals with better coverages than those produced by the homoscedastic models.

	RMSFE			MAD		
Local Level	$h=1$	$h=6$	$h=12$	$h=1$	$h=6$	$h=12$
Homoscedastic w/o breaks	0.920	0.925	1.017	0.959	0.851	0.956
Homoscedastic w/ breaks	0.890*	0.882**	0.946*	0.919*	0.825**	0.897*
GARCH in ε_t w/o breaks	0.905*	0.898*	0.979	0.939	0.836*	0.939
GARCH in ε_t w/ breaks	0.877*	0.862**	0.933	0.897*	0.803**	0.895*
IGARCH in both noises	0.906	0.902*	0.984	0.936	0.834	0.930*
ARSV in ε_t and RW-SV in η_t	0.918	0.913	1.006	0.954	0.834*	0.954
IMA(1,1)						
Homoscedastic	0.914	0.916	1.004	0.949	0.844	0.945
GARCH in a_t	0.914	0.916	1.004	0.949	0.844	0.945
IGARCH in a_t	0.912	0.915	1.003	0.938	0.844*	0.945
Rolling Window TV-IMA(1,1)						
S & W approach	1.000	1.000	1.000	1.000	1.000	1.000
GARCH in a_t	0.998	0.990*	1.000	0.995	0.986*	1.004
IGARCH in a_t	0.909	0.894**	0.972	0.926	0.860*	0.950

Table 4.5: Out-of-sample evaluation of the point forecasts for the inflation series, from 6/2003 to 5/2008. RMSFE is the root mean squared forecast error and MAD the mean absolute deviation. The values are standardized for each horizon h, such that the RMSFE and MAD of the benchmark model labelled as *S & W approach* are equal to one. *(**) means 5%(1%) significantly more accurate than the benchmark, according to the Diebold and Mariano (1995) test.

4.4 Conclusions

In this chapter, we study the use of a rolling-windows approach to estimate the reduced form MA parameter, θ, when the series is generated by the local level model with different specifications in its noises. With some simulated series, we observe that if the marginal variances of ε_t and η_t are constant, then θ is also constant over time, regardless of whether these noises are heteroscedastic or not. On the other hand, when assuming integrated processes on the variances of the two noises, θ is not defined and therefore we cannot say anything about its evolution. Finally, when the marginal variances have one or more breaks, this is going to affect θ. Consequently, with these results, we say that the rolling window approach considered in Stock and Watson (2007), cannot be used to track the evolution of the stationary conditional variances of ε_t and η_t, and in general, that it is very difficult to assess if the evolution of the $\hat{\theta}_{rol}$ comes only from sample variability or from changes in the marginal variances of the local level noises. Moreover, we also show that the MA model with a rolling window estimate of θ is not able to capture the conditionally heteroscedastic structure of the local level noises.

From the results of the forecasting exercise with the selected inflation series, we conclude that the conditionally heteroscedastic models produce more accurate point and interval forecast than those with homoscedastic noises. Moreover, the results on the point forecast evaluation suggest that the conditionally heteroscedastic local level model is in general better than the

	90%			95%		
Local Level	$h=1$	$h=6$	$h=12$	$h=1$	$h=6$	$h=12$
Homoscedastic w/o breaks	83.33	85.45	89.80	91.67	92.73	91.84
Homoscedastic w/ breaks	83.33	81.82	83.67	90.00	89.09	89.80
GARCH in ε_t w/o breaks	91.67	90.91	93.88	93.33	96.36	93.88
GARCH in ε_t w/ breaks	85.00	87.27	89.80	91.67	90.91	91.84
IGARCH in both noises	91.67	94.55	93.88	96.67	96.36	93.88
ARSV in ε_t and RW-SV in η_t	93.33	89.09	91.84	93.33	96.36	93.88
IMA(1,1)						
Homoscedastic	83.33	85.45	89.80	91.67	92.73	91.84
GARCH in a_t	93.33	94.55	93.88	95.00	96.36	93.88
IGARCH in a_t	91.67	94.55	93.88	96.67	96.36	95.92
Rolling Window TV-IMA(1,1)						
S & W approach	86.67	81.82	81.63	90.00	89.09	87.76
GARCH in a_t	88.33	90.91	85.71	91.67	92.73	93.88
IGARCH in a_t	90.00	89.09	87.76	93.33	92.73	93.88

Table 4.6: Out-of-sample empirical coverage of the 90% and 95% prediction intervals for the inflation series, from 6/2003 to 5/2008. The observed coverage is defined as the proportion of times the actual value of the series lays within the 90% and 95% prediction intervals, for each horizon h.

IMA counterparts, and that the inclusion of the deterministic break in the marginal variance of η_t significantly improves forecast accuracy, for almost all series and horizons. However, in what concerns to interval forecasting evaluation, results are no so clear, and there is not a dominant model for all series. In any case, the models with conditionally heteroscedastic noises seem to generate prediction intervals with coverages closer to the nominal.

Chapter 5

On the accuracy of prediction intervals when using a disaggregation approach

5.1 Introduction

The effect of disaggregation on the forecasts of index series has been widely analyzed in the literature; see Espasa and Albacete (2004), Hubrich (2005) and Hendry and Hubrich (2006) for a detailed study of this topic and applications for the Harmonized Consumer Price Index of the Euro zone. Within the context of vector ARMA (VARMA) models, forecasting contemporaneous and temporal aggregations of different series has been widely analyzed since the seminal work of Lütkepohl (1987). On the other hand, aggregation in conditionally heteroscedastic models in the form of GARCH processes has also been considered in the literature; see Silvestrini and Veredas (2008) for an up-to-date survey on aggregation of VARMA models with multivariate GARCH noises.

In this chapter, we study the performance of the prediction intervals of the aggregate when at least one of its components is conditionally heteroscedastic. In particular, we analyze if using the indirect approach of modelling the whole vector of components improves the accuracy of prediction intervals when compared with the direct approach of taking the aggregate as a univariate series. We obtain the predictions of the conditional means and variances k steps ahead for the case of a VAR(p) process with a constant conditional correlation (CCC) GARCH residual vector. We show with simulated data that the indirect approach may be preferred because it yields more accurate prediction intervals than working directly with the aggregate.

The rest of the chapter is structured as follows. In Section 5.2 we briefly review the main findings on aggregation of GARCH processes. In Section 5.3, we first derive the multi-step minimum mean squared error (MSE) prediction of the conditional means and variances of the VAR(p) process with CCC-GARCH noises. Next, we find the prediction intervals of the aggregate, using either its own past or the information coming from the multivariate process,

and particularize the results for the bivariate VAR(1) model with mutually independent noises. Section 5.4 presents several Monte Carlo simulations to study the relative accuracy of the two alternative ways of constructing the prediction intervals of the aggregate. Finally, Section 5.5 concludes.

5.2 Aggregation of GARCH models: A brief review

In these days, it is almost indisputable that many financial and monetary time series such as stock returns, exchange rates, or price indexes exhibit conditional heteroscedasticity. Among all the existing parametric models that account for this feature, those of the GARCH family, originally proposed by Engle (1982) and Bollerslev (1986), are the most popular, specially in empirical applications. A recent review of the applications of GARCH processes is given by Diebold (2004).

In its simplest form, the GARCH(p,q) process is given by

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \, \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \, \sigma_{t-j}^2, \tag{5.1}$$

where ε_t is a sequence of stationary errors and the coefficients ω, α_i, $i = 1, .., q$, and β_j, $j = 1, 2, ..., p$, are such that the fourth order moment of ε_t is finite and σ_t^2 is nonnegative.

One of the theoretical drawbacks of GARCH models is that they are not necessarily closed under aggregation. That is, the sum of two or more GARCH processes is not in general GARCH. With respect to temporal aggregation, Drost and Nijman (1993) conclude that only weaker versions of the GARCH models are closed under temporal aggregation. They define three types of GARCH models, strong, semi-strong and weak, with the following properties:

$$\varepsilon_t^{\dagger} = \varepsilon_t/\sigma_t \sim i.i.d.\, D(0,1) \quad \text{(Strong definition)} \tag{5.2}$$

$$E[\varepsilon_t^i | \varepsilon_{t-1}, \varepsilon_{t-2}, ...] = \begin{cases} 0, & i = 1 \\ \sigma_t^2, & i = 2 \end{cases} \quad \text{(Semi-Strong definition)} \tag{5.3}$$

$$P[\varepsilon_t^i | \varepsilon_{t-1}, \varepsilon_{t-2}, ...] = \begin{cases} 0, & i = 1 \\ \sigma_t^2, & i = 2 \end{cases} \quad \text{(Weak definition)} \tag{5.4}$$

where $D(0,1)$ specifies a distribution with zero mean and unit variance, and $P[x_t|\varepsilon_{t-1}, \varepsilon_{t-2}, ...]$ denotes the best linear predictor of x_t in terms of $1, \varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, ...$, i.e

$$E\left(x_t - P[x_t|\varepsilon_{t-1}, \varepsilon_{t-2}, ...]\right) \varepsilon_{t-i}^r = 0, \; i \geq 1, \; r = 0, 1, 2. \tag{5.5}$$

Note that the weak definition implies that ε_t must be a zero-mean uncorrelated process in which $Cov(\varepsilon_t^2, \varepsilon_{t-i}^2) = Cov(\sigma_t^2, \sigma_{t-i}^2)$, $i \geq 1$. Drost and Nijman (1993) show that even in this general version of GARCH processes, it is possible to obtain strongly consistent QML estimators. They also show that the class of ARMA models with symmetric weak GARCH errors and either stock or flow variables are closed under temporal aggregation. In a more recent paper, Meddahi and Renault (2004) discuss the results found by Drost and Nijman (1993) and state that the

limitations of the weak GARCH processes, specially when trying to capture the empirical features of the series, make them less attractive and propose a new model based on linear state space modelling (or stochastic volatility modelling), that is also closed under temporal aggregation.

On the other hand, Nijman and Sentana (1996) study the behavior of GARCH models under contemporaneous aggregation. They show that the sum of independent univariate GARCH processes yields a weak GARCH process. In particular, they demonstrate that the sum of two independent (strong) GARCH(1,1) process yields a weak GARCH(2,2) process. There are however two special cases, which consist on adding a mutually independent white noise error to a (strong) GARCH(1,1) process and summing two (strong) GARCH(1,1) process with the same persistence[1]. In these two cases, the result is another (weak) GARCH(1,1) process. In the limiting case, when the aggregation involves a large number of GARCH processes, Zaffaroni (2007) shows that the heteroscedasticity in the limiting aggregate depends crucially on the cross correlations among the components. Thus, for instance, the limiting aggregate of n GARCH processes, as n goes to infinity, maintains the features of a volatility model (uncorrelated levels and correlated squares) only for sufficiently cross-correlated components and has a linear stable representation otherwise. For financial time series aggregates like stock indexes, the results of Zaffaroni (2007) are relevant because even with mutually independent stock returns, there exist situations in which the volatility is not diversified away by aggregation.

5.3 Prediction intervals of the aggregate

In this section, we obtain the prediction intervals of the aggregate, y_t, which is composed of n possibly conditionally heteroscedastic components, $x_{i,t}$, stacked in a vector X_t. We assume that X_t follows a VAR(p) process with contemporaneously correlated noises. As the interest lays in studying prediction intervals, we need first to find the multi-step minimum mean squared error (MSE) prediction of the aggregate and its conditional variance. Throughout this section we will assume that the parameters are known unless otherwise stated.

5.3.1 The VAR(p) with Multivariate GARCH disturbances

Consider that X_t is a vector of n stationary time series used to construct an aggregate y_t. Assume that the dynamics of X_t is represented by the following VAR(p) model

$$X_t = C + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + ... + \Phi_p X_{t-p} + \varepsilon_t, \tag{5.6}$$

where C is a vector of constants and Φ_i, $i = 1, ..., p$ are $n \times n$ matrices of coefficients such that all the roots of

$$\Phi(L) = I_n - \Phi_1 L - \Phi_2 L^2 - ... - \Phi_p L^p,$$

[1] In a GARCH(1,1) process, the literature often defines the measure of persistence as the sum of the ARCH and GARCH coefficients, i.e. $\alpha_1 + \beta_1$.

lay outside the unit circle. Consider also that ε_t has the following properties

$$\begin{aligned} \varepsilon_t &= \Sigma_t^{1/2} Z_t, \\ & Z_t \sim NID_n(\mathbf{0}, R), \\ & \Sigma_t^{1/2} = \text{diag}(\sigma_{1,t}, ..., \sigma_{n,t}), \end{aligned} \tag{5.7}$$

where NID_n denotes the n^{th}-dimensional serially independent Normal distribution, $R = [\rho_{i,j}]$, $\rho_{i,i} = 1$ is the covariance matrix of Z_t. Additionally,

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2, \ i = 1, 2, ..., n, \tag{5.8}$$

with ω_i, α_i and β_i chosen to satisfy fourth order stationarity and positiveness of $\sigma_{i,t}^2$. This model for ε_t is a simple multivariate version of the GARCH(1,1), allowing for constant conditional correlation (CCC) among the errors, $\varepsilon_{i,t}$, and it was first proposed by Bollerslev (1990); see Ling and McAleer (2003) and He and Terasvirta (2004) for extensions and generalizations. Expression (5.7) implies that $\varepsilon_{i,t}$ are serially uncorrelated with constant contemporaneous cross correlations. Moreover, the lagged cross covariances, $E[\varepsilon_{i,t}, \varepsilon_{j,t-s}]$, with $s > 0$, are also zero for all $i \neq j$. With (5.6) and (5.7), we say that X_t follows a VAR(p) process with CCC-GARCH noises. Using the properties of this model, we know that the mean and variance of X_{t+1}, conditional on the information up to time t are given by

$$\begin{aligned} X_{t+1|t} &\equiv \underset{X_t}{E}[X_{t+1}] \\ &= C + \Phi_1 X_t + \Phi_2 X_{t-1} + ... + \Phi_p X_{t-p+1} \end{aligned} \tag{5.9}$$

$$\begin{aligned} V_{t+1|t} &\equiv \underset{X_t}{E}[(X_{t+1} - X_{t+1|t})(X_{t+1} - X_{t+1|t})'] \\ &= \underset{X_t}{E}[\Sigma_{t+1}^{1/2} Z_{t+1} Z_{t+1}' \Sigma_{t+1}^{1/2}] \\ &= \Sigma_{t+1}^{1/2} R \Sigma_{t+1}^{1/2}, \end{aligned} \tag{5.10}$$

where $\underset{X_t}{E}$ denotes the mean conditional on the whole set $\{X_1, X_2, ..., X_t\}$. Note that the since Z_t is Normal, then X_{t+1} given X_t, i.e. $X_{t+1}|X_t$, is also Normal, with mean $X_{t+1|t}$ and variance $V_{t+1|t}$. However, for multi-step forecasts, i.e. $k > 1$, due to the multiplicative structure of the CCC-GARCH model, $X_{t+k}|X_t$ is not Normal anymore. Additionally, we do not have an analytic expression for the k-step-ahead conditional variance of ε_t, defined as $H_{t+k|t} \equiv \underset{X_t}{E}[\varepsilon_{t+k} \varepsilon_{t+k}']$, unless R is diagonal, i.e. $\rho_{i,j} = 0$ for all $i \neq j$. In this case, we find that

$$\begin{aligned} H_{t+k|t} &= \Sigma_{t+k|t} \\ &= \text{diag}\left(\sigma_{1,t+k|t}^2, ..., \sigma_{n,t+k|t}^2\right), \end{aligned} \tag{5.11}$$

where, by property of univariate GARCH(1,1) processes,

$$\begin{aligned} \sigma_{i,t+k|t}^2 &\equiv \underset{\sigma_{i,t}^2}{E}[\sigma_{i,t+k}^2] \\ &= \sigma_i^2 + (\alpha_i + \beta_i)^{k-1}(\sigma_{i,t+1}^2 - \sigma_i^2), \ k > 1, \end{aligned} \tag{5.12}$$

for $i = 1, 2, ..., n$, and σ_i^2 is the unconditional variance of $\varepsilon_{i,t}$. For the general case in which R is not diagonal, the characteristic element of $H_{t+k|t}$ is given by $h_{i,j,t+k|t} = \rho_{i,j} \underset{X_t}{E}[\sigma_{i,t+k} \sigma_{j,t+k}]$,

which is equal to $\sigma^2_{i,t+k|t}$ when $i = j$. The conditional mean included in $h_{i,j,t+k|t}$ does not have an analytic expression as a function of the GARCH parameters. However, we can approximate the off-diagonal elements of $H_{t+k|t}$ by assuming that $\underset{X_t}{E}[\sigma_{i,t+k}\sigma_{i,t+k}] \approx \sigma_{i,t+k|t}\,\sigma_{j,t+k|t}$. In this case, if we believe that this approximation is valid, we can obtain the multi-step conditional variance of ε_t given by

$$H_{t+k|t} \approx \Sigma^{1/2}_{t+k|t}\, R\, \Sigma^{1/2}_{t+k|t}, \tag{5.13}$$

We find now the multi-step predictions of X_t. Applying conditional expectations to (5.6), the minimum MSE k-step ahead predictor for X_{t+k}, $X_{t+k|t}$, is given by

$$\begin{aligned} X_{t+k|t} &\equiv \underset{X_t}{E}[X_{t+k}] \\ &= C + \sum_{j=1}^{p} \Phi_j\, X_{t+k-j|t}, \end{aligned} \tag{5.14}$$

where $X_{t+k-j|t} = X_{t+k-j}$ when $j \geq k$. Therefore, we can work recursively by obtaining $X_{t+j|t}$ as a function of $X_{t+j-1|t}$ and use the result to obtain $X_{t+j+1|t}$ until $j + 1 = k$. On the other hand, having found $H_{t+k|t}$, we can also find the conditional variance of the k-steps ahead MSE point predictor of X_{t+k}, $V_{t+k|t}$. First, we follow Hlouskova et al. (2009) and put model (5.6) in the following companion format

$$\overline{X}_t = \overline{C} + \Omega\, \overline{X}_{t-1} + \overline{\varepsilon}_t, \tag{5.15}$$

where

$$\overline{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix}, \quad \overline{C} = \begin{bmatrix} C \\ 0_n \\ \vdots \\ 0_n \end{bmatrix}, \quad \Omega = \begin{bmatrix} \Phi_1 & & \dots & & \Phi_p \\ I_n & 0_n & \dots & & 0_n \\ \dots & & \dots & & \dots \\ 0_n & & \dots & I_n & 0_n \end{bmatrix}, \quad \overline{\varepsilon}_t = \begin{bmatrix} \varepsilon_t \\ 0_n \\ \vdots \\ 0_n \end{bmatrix},$$

and I_n and 0_n are the $n \times n$ identity and null matrices, respectively. This companion format is useful because now we work with a (higher dimensional) VAR(1). By property of this model, we know that the conditional variance of X_{t+k} is given by

$$\begin{aligned} V_{t+k|t} &\equiv \underset{X_t}{E}\left[(X_{t+k} - X_{t+k|t})(X_{t+k} - X_{t+k|t})'\right] \\ &= B' \sum_{j=0}^{k-1} \Omega^j\, B\, H_{t+k-j|t}\, \left(\Omega^j\, B\right)' B, \end{aligned} \tag{5.16}$$

where $B = [I_n, 0_n, ..., 0_n]'$ is an $np \times n$ matrix used to extract the first n elements of $\overline{X}_t$, i.e. the ones corresponding to X_t. Note from (5.16) that, as long as $H_{t+k|t}$ is approximated by (5.13), $V_{t+k|t}$ is also an approximation of the true conditional variance.

5.3.2 Prediction intervals of the aggregate

Consider the $n \times 1$ weighting vector, λ_t, that is known by the forecaster at the time the prediction is made[2], such that the aggregate y_t is given by

$$\begin{aligned} y_t &= \lambda_t' X_t \\ &= \sum_{i=1}^{n} \lambda_{i,t}\, x_{i,t}, \end{aligned} \tag{5.17}$$

with X_t defined as in (5.6). Then, y_t is a weighted sum of the elements of a stationary VAR(p) process. Lütkepohl (1984) shows that in this case, the aggregation of such process is an ARMA(p',q') model of the form

$$y_t = c + \sum_{i=1}^{p'} \phi_i\, y_{t-i} + \sum_{j=1}^{q'} \theta_j\, a_{t-j} + a_t, \tag{5.18}$$

where a_t is a zero-mean uncorrelated process, and c, ϕ_i and θ_j parameters such that the process is stationary and invertible. This means that the stationary VAR(p) process on X_t is closed to contemporaneous aggregation. Then, given the orders p' and q', we can use this univariate model to forecast y_t, conditional on its own past, i.e the whole set $\{y_1,\ y_2, ..., y_t\}$. The minimum MSE point predictor of y_{t+k}, according to (5.18), is given by

$$\begin{aligned} y_{t+k|t} &\equiv \underset{y_t}{E}[y_{t+k}] \\ &= c + \sum_{i=1}^{p'} \phi_i\, y_{t+k-i|t} + \sum_{j=1}^{q'} \theta_j\, a_{t+k-j}, \end{aligned} \tag{5.19}$$

where $y_{t+k-i|t} = y_{t+k-i}$ when $i \geq k$ and $a_{t+k-j} = 0$ when $j \geq k$. In (5.19), we assume that, at time t, the whole series of innovations $\{a_1, ..., a_t\}$ can be recovered once we know c, ϕ_i and θ_j.

Alternatively, we can work with $X_{t+k|t}$ given in (5.14) and directly aggregate the MSE point prediction of X_t, that is

$$y^*_{t+k|t} = \lambda_t' X_{t+k|t}. \tag{5.20}$$

When comparing $y_{t+k|t}$ with $y^*_{t+k|t}$, the main results found in the literature state that if the data generating process (DGP) of the vector X_t is known in terms of structure and coefficients, working with the indirect approach of first obtaining $X_{t+k|t}$ and then calculating $y^*_{t+k|t}$ is preferred in terms of smaller mean square forecast errors typically because the information set in this case, the whole set of past values of X_t, is greater than the information set composed of only the past values of y_t; see Lütkepohl (1987) and Hendry and Hubrich (2006) for a deeper study. However, when the DGP is unknown, even if the order selection criterion is consistent, the point forecasts coming from the indirect approach may not always be more accurate than those coming from the direct approach of using the past values of y_t. In fact, as Hubrich (2005)

[2]Note that the weights may be time-varying, a feature that is commonly observed in economic aggregates like the consumer price index (CPI).

states, the relative (point) forecast accuracy depends on the extent to which the gains derived from a greater information set are offset by the effects of estimation variability.

When analyzing the prediction intervals of both alternative ways of forecasting the aggregate, y_t, we find a big difference with respect to point forecasts. As discussed in the previous section, if we assume a certain type of multivariate GARCH process on the residuals ε_t, when aggregating this vector the result does not belong to any class of (strong) univariate GARCH. Indeed, even in the simplest case where R is diagonal, we obtain the result of Nijman and Sentana (1996) with independent univariate GARCH processes, generalized for the case of n components. Therefore, we cannot explicitly characterize the volatility dynamics of a_t with a (strong) GARCH(p,q) model. However, we do know some properties of this process. For example, we know that in general, the heteroscedastic structure of a_t is weaker than that of the components ε_{it}, something that has been studied in Chapter 2 but in a different context. Therefore, often one cannot reject the null of homoscedasticity in the aggregate that is composed of one or more conditionally heteroscedastic components. Moreover, we also know that if the GARCH processes on ε_{it} share some properties, then their contemporaneous aggregation still provides the best linear predictor of the variance, i.e. a_t follows a weak GARCH process. In any case, if we define $h_{t+k|t} \equiv \underset{y_t}{E}[a_{t+k}^2]$ as the k-step ahead conditional variance of a_{t+k} in (5.18), then the conditional variance of $y_{t+k|t}$, according to model (5.18)$v_{t+k|t}$, is given by

$$\begin{aligned} v_{t+k|t} &\equiv \underset{y_t}{E}\left[(y_{t+k} - y_{t+k|t})^2\right] \\ &= b' \sum_{j=0}^{k-1} \Theta^j \, b_1 \, h_{t+k-j|t} \, \left(\Theta^j \, b_1\right)' b, \end{aligned} \tag{5.21}$$

where

$$\Theta = \begin{bmatrix} \phi_1 & & \dots & & \phi_{p'} & \theta_1 & \dots & & \theta_{q'} \\ 1 & 0 & \dots & & 0 & 0 & \dots & & 0 \\ \dots & & \dots & & \dots & \dots & \dots & & \dots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & & 0 \\ 0 & 0 & \dots & 1 & 0 & 1 & \dots & & 0 \\ \dots & & \dots & & \dots & \dots & \dots & & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

is the $p' + q'$ square matrix of coefficients of the VAR(1) model applied to the stacked vector

$$\overline{y}_t = [y_t, y_{t-1}, \dots, y_{t-p'}, a_t, a_{t-1}, \dots, a_{t-q'}]',$$

$b_1 = [1, 0, ..., 0, 1]'$ and $b = [1, 0, ..., 0]'$ are two $(p' + q') \times 1$ vectors used to extract the first element of $\overline{y}_t$. With $y_{t+k|t}$ and $v_{t+k|t}$, we are able to construct the $(1 - \alpha)$ prediction interval k steps ahead, $PI_k(1 - \alpha)$ as

$$PI_k(1-\alpha) = \left[\mathcal{F}_{\alpha/2}(y_{t+k|t}, v_{t+k|t}) \, ; \, \mathcal{F}_{1-\alpha/2}(y_{t+k|t}, v_{t+k|t})\right], \; k \geq 1, \tag{5.22}$$

where $\mathcal{F}$ is the cumulative distribution function of $y_{t+k}|y_t$, with mean $y_{t+k|t}$ and variance $v_{t+k|t}$. Note that a_t being and aggregation of GARCH processes implies that $\mathcal{F}$ is not Normal and,

in general, we need to use, for example, bootstrap methods to estimate $\mathcal{F}$; see Pascual et al. (2006) for a deep study. However, as these authors show, Gaussian intervals may be a good approximation to the true intervals, so that we can approximate (5.22) as

$$PI_k(1-\alpha) \approx y_{t+k|t} \pm \mathcal{N}_{\alpha/2}^{-1} \, v_{t+k|t}^{1/2}, \; k \geq 1, \tag{5.23}$$

where $\mathcal{N}_u^{-1}$ is the argument of the standard Normal cumulative distribution function valued at u.

Alternatively, we can work directly with $X_{t+k|t}$ given in (5.14) and find the conditional variance of $y_{t+k|t}^*$. Then, it is straightforward to see that

$$v_{t+k|t}^* = \lambda_t' \, V_{t+k|t} \, \lambda_t. \tag{5.24}$$

Using an analogous approximation to (5.23), we find the $(1-\alpha)$ prediction interval k steps ahead, $PI_k^*(1-\alpha)$ as

$$PI_k^*(1-\alpha) \approx y_{t+k|t}^* \pm \mathcal{N}_{\alpha/2}^{-1} \, (v_{t+k|t}^*)^{1/2}, \; k \geq 1. \tag{5.25}$$

As with the point predictions, we are interested in studying the accuracy of the two prediction intervals, PI_k and PI_k^*, in terms of their coverage. In order to clarify the analysis, we will focus on a simple bivariate VAR(1) model. In the next paragraphs, we find closed form expressions of the conditional mean and variance of both approaches for this simple case.

5.3.3 The bivariate VAR(1) model with independent GARCH noises

Let X_t be a 2×1 vector that follows a VAR(1) process with zero mean, i.e. $C = 0$, and with the noises $\varepsilon_{i,t}$ being serially and mutually uncorrelated, each of them following a univariate GARCH(1,1) process. This is the simplest version of the general n^{th} dimensional VAR(p) model with CCC-GARCH noises. It can be shown that the aggregate of X_t in this particular case is an ARMA(2,1). Of course, different restrictions in the parameters space of the VAR(1) may lead to an ARMA process of a lower order, i.e. some roots of the AR and MA polynomials cancel out.

With respect to the conditional mean of X_t, expression (5.14) reduces to

$$\begin{aligned} X_{t+k|t} &= \Phi \, X_{t+k-1|t} \\ &= \Phi^k \, X_t, \end{aligned} \tag{5.26}$$

where Φ is the matrix of coefficients of the VAR(1). On the other hand, using the result of (5.11), the conditional variance of X_t is given by

$$V_{t+k|t} = \sum_{j=0}^{k-1} \Phi^j \, \Sigma_{t+k-j|t} \, \left(\Phi^j\right)'. \tag{5.27}$$

Note that the marginal variance of X_t, $V[X_t]$, is obtained by substituting $\Sigma_{t+k|t}$ for the marginal variance of the errors, $\Sigma_\varepsilon = \text{diag}(\sigma_1^2, \; \sigma_2^2)$. Then, by means of (5.12), we can obtain a multivariate

version of the excess volatility in the following way:

$$V_{t+k|t} - V[X_t] = \sum_{j=0}^{k-1} \Phi^j \operatorname{diag}\left[(\alpha_1+\beta_1)^{k-1}(\sigma^2_{1,t+1}-\sigma^2_1),\ (\alpha_2+\beta_2)^{k-1}(\sigma^2_{2,t+1}-\sigma^2_2)\right] \left(\Phi^j\right)'. \tag{5.28}$$

Note from expression (5.28) that the excess volatility of the component $x_{i,t}$ one step ahead is simply that of its corresponding noise, $\varepsilon_{i,t}$. However, as k increases, the excess volatility of $x_{i,t}$ depends not only on the excess volatility of $\varepsilon_{i,t}$, but also on that of $\varepsilon_{j,t}$, with $i = 1, 2$ and $i \neq j$, by means of matrix Φ. This means that independent GARCH noises on the VAR process does not ensure independent paths for the conditional variances of each component[3]. This is an important result because it states that if one component is homoscedastic, its conditional variance may still be influenced by periods of high or low volatility of other components.

When aggregating X_t by means of a vector $\lambda = [\lambda_1,\ (1-\lambda_1)]'$ (we suppress the time index for simplicity), $y^*_{t+k|t}$ and $v^*_{t+k|t}$ are given by

$$y^*_{t+k|t} = \lambda' \Phi^j X_t \tag{5.29}$$

$$v^*_{t+k|t} = \lambda' \sum_{j=0}^{k-1} \Phi^j \Sigma_{t+k-j|t} \left(\Phi^j\right)' \lambda. \tag{5.30}$$

In this case, the excess volatility of the aggregate when $k = 1$ is equal to

$$v^*_{t+1|t} = \lambda_1^2(\sigma^2_{1,t+1|t} - \sigma^2_1) + (1-\lambda_1)^2(\sigma^2_{2,t+1|t} - \sigma^2_2).$$

That is, a combination of the excess volatilities of $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$. Note that if $0 < \lambda_1 < 1$, the excess volatility of the aggregate may be closer to zero than that of all the components, because $\lambda_1^2 + (1-\lambda_1)^2 < 1$. Finally, we obtain the approximated $(1-\alpha)$ prediction intervals of y_t conditional on the information given by X_t directly by plugging (5.29) and (5.30) into (5.25).

Alternatively, if we take y_t and find the ARMA(p',q') process that results of aggregating X_t, it is relatively easy to see that $p' = 2$ and $q' = 1$, that is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_1 a_{t-1} + a_t. \tag{5.31}$$

Moreover, following the development of Hendry and Hubrich (2006), we know that $\phi_1 = \operatorname{trace}(\Phi)$, $\phi_2 = -\det(\Phi)$ and θ_1 and σ_a^2 are the parameters of the MA(1) process given by

$$m_t = \lambda_1 (\varepsilon_{1,t} - u_1 \varepsilon_{1,t-1}) + (1-\lambda_1)(\varepsilon_{2,t} - u_2 \varepsilon_{2,t-1}), \tag{5.32}$$

where $u_1 = \Phi_{2,2} - \Phi_{2,1}$ and $u_2 = \Phi_{1,1} - \Phi_{1,2}$, with $\Phi_{i,j}$ being the element of the i^{th} row and j^{th} column of Φ. Consequently, we are able to fully characterize the conditional mean of the aggregate process. However, as we mentioned above, a_t is not in general GARCH, so that it is impossible to find the conditional variance by doing something similar to the conditional mean.

[3]Only in the case in which Φ is diagonal, i.e. no component causes in the Granger sense to the others, we can say that the excess volatility in $\varepsilon_{i,t}$ affects the conditional variance of only $x_{i,t+k}$. Obviously, this is the case of two totally independent components, that can be modelled separately.

In any case, we may use some of the results given by Nijman and Sentana (1996) commented before, and define a (weak) GARCH process for a_t. Thus, once we fit a model for the conditional mean and another for the conditional variance, we are able to plug the point forecast and its conditional variance into the prediction interval of (5.23).

5.3.4 Cumulative forecasts and their conditional variance

Finally, consider now that X_t is the stationary transformation of a given time series vector P_t such that $X_t = \Delta P_t$. This may be the case of a (log) price index, in which the inflation process, defined as its first difference, is given by X_t. Then, it may be interesting to find the point and interval forecasts directly of P_t instead of X_t. In this case, under the assumptions made on X_t, it is easy to show that the k-steps ahead MSE point predictor, $P_{t+k|t}$ is given by

$$P_{t+k|t} = \sum_{j=1}^{k} X_{t+k|t},\ k \geq 1. \tag{5.33}$$

Let $W_{t+k|t}$ be the minimum MSE conditional variance of P_{t+k}. It is straightforward to see that $W_{t+1|t}$ is equal to $H_{t+1|t}$, which is also equal to $V_{t+1|t}$. However, for $k > 1$, $W_{t+k|t}$ has a more complex expression because it not only involves the sum of the conditional variances $V_{t+k|t}$, but also of the covariances between the forecasts i and j periods ahead, which are not in general zero. After working out the conditional expectations involved in $W_{t+k|t}$, we obtain a recursive expression as follows

$$W_{t+k|t} = W_{t+k-1|t} + V_{t+k|t} + \Phi\, K_\Phi\, V_\Phi + (K_\Phi\, V_\Phi)'\, \Phi',\ k > 1, \tag{5.34}$$

where

$$K_\Phi = [I_2,\ \Phi,\ \Phi^2,\ \ldots,\ \Phi^{j-2}]$$

and

$$V_\Phi = [V_{t+k-1|t}, V_{t+k-2|t},\ \ldots,\ V_{t+1|t}]'$$

are two stacked matrices used to obtain the cross products involved in the covariances. With expressions (5.33) and (5.3.4) we find the cumulative forecasts and their conditional variance of the aggregate by noting that $p^*_{t+k|t} = \lambda'\, P_{t+k|t}$, so that $w^*_{t+k|t} = \lambda'\, W_{t+k|t}\, \lambda$.

On the other hand, the multi-step cumulative forecasts of the univariate ARMA(2,1) process for the aggregate y_t is given by

$$p_{t+k|t} = \sum_{j=1}^{k} y_{t+k|t},\ k \geq 1, \tag{5.35}$$

with their conditional variance given by where

$$w_{t+k|t} = b'\, \Psi_{t+k|t}\, b,\ k > 1, \tag{5.36}$$

where

$$\Psi_{t+k|t} = \Psi_{t+k-1|t} + \Xi_{t+k|t} + \Theta\, K_\Theta\, \Xi_\Theta + (K_\Theta\, \Xi_\Theta)'\, \Theta',$$

$$K_\Theta = [I_3,\ \Theta,\ \Theta^2,\ \ldots,\ \Theta^{j-2}],$$

$$\Xi_\Theta = [\Xi_{t+k-1|t},\ \Xi_{t+k-2|t},\ \ldots,\ \Xi_{t+1|t}]',$$

$$\Xi_{t+k|t} = \sum_{j=0}^{k-1} \Theta^j\, b_1\, h_{t+k-j|t}\, \left(\Theta^j\, b_1\right)'$$

and $\Psi_{t+1|t} = \Xi_{t+1|t}$. The vectors b and b_1 are defined as in (5.21). With $p^*_{t+k|t}$, $w^*_{t+k|t}$, $p_{t+k|t}$ and $w_{t+k|t}$, we can obtain the approximated k-steps-ahead prediction intervals given in (5.25) and (5.23), respectively. Note that the cumulative forecasts of y_t, using either the indirect or direct approach, have conditional variances that increase with the horizon k. This also implies that the effects of a volatility shock at the time of forecasting do not vanish with k.

In the next section, we perform several Monte Carlo exercises with data simulated from the VAR(1) process with CCC-GARCH noises to analyze the accuracy of the prediction intervals generated by either the direct approach of using y_t as a univariate process or the indirect approach of using the information of the whole vector X_t. We also perform the exercise for the cumulative forecasts using the expressions given above.

5.4 A Monte Carlo exercise

The objective in this section is to evaluate the coverage of the prediction intervals for different specifications of the VAR(p) model with CCC-GARCH noises, given in (5.6). We stick to the bivariate case and $p = 1$, allowing for GARCH(1,1) noises in either or both components. Table (5.1) describes the data generating processes used to simulate the data. Note that the range of models cover the simplest cases of R and/or Φ being diagonal and the more complex ones where the unrestricted VAR(1) is combined with the mutually dependent CCC-GARCH, i.e. R is also unrestricted. Without loss of generality, we use the equally weighted index that implies $\lambda_1 = 0.5$. We also fix

$$\Phi = \begin{bmatrix} 0.4 & -0.3 \\ 0.5 & 0.8 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}.$$

Additionally, we set $\alpha_1 = \alpha_2 = 0.15$, $\beta_1 = \beta_2 = 0.8$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 2$ for the univariate GARCH processes for the unrestricted versions of the models. In the restricted cases, we just fix to zero the corresponding elements of the two matrices, R and Φ, and set $\alpha_1 = \beta_1 = 0$ for the cases in which only the second component is heteroscedastic.

For each DGP, we generate $B = 1000$ bivariate processes X_t and, assuming that the parameters are known, we find the MSE point predictor of X_{t+k} given X_t and its conditional variance using the expressions given above. Then, we use them to obtain the prediction intervals of the aggregate. This will be the indirect approach, which counts with the information on X_t. Alternatively, we obtain the univariate process y_t of the simulated X_t, by means of the relationships between the parameters given above, and find the point forecasts. With respect to the conditional variance, we use the corresponding (weak) GARCH process[4]. This exercise allows us to

[4]For the more general cases included in the DGPs of Table 5.1, for which we do not count with a theoretical

	Conditional mean		Conditional Variance	
Label	Order (p)	Matrix Φ	GARCH(1,1)	Matrix R
U-B-D	1	Upper triangular	Both	Diagonal
F-B-D	1	Full	Both	Diagonal
D-B-D	1	Diagonal	Both	Diagonal
U-2^{nd}-D	1	Upper triangular	2^{nd} Component	Diagonal
F-2^{nd}-D	1	Full	2^{nd} Component	Diagonal
D-2^{nd}-D	1	Diagonal	2^{nd} Component	Diagonal
U-B-F	1	Upper triangular	Both	Full
F-B-F	1	Full	Both	Full
D-B-F	1	Diagonal	Both	Full
U-2^{nd}-F	1	Upper triangular	2^{nd} Component	Full
F-2^{nd}-F	1	Full	2^{nd} Component	Full
D-2^{nd}-F	1	Diagonal	2^{nd} Component	Full

Table 5.1: Data generating processes (DGPs) used for the Monte Carlo experiments. The matrix of coefficients of the VAR(1) model, Φ, may be unrestricted (Full), upper triangular or diagonal. The error vector ε_t may have GARCH(1,1) processes in only the second or in both components, allowing for constant conditional correlation different from zero (R is Full) or zero (R is Diagonal).

compare the coverage of the two prediction intervals when we know the DGP of the series. Alternatively, it may be worthy to fit also the best VAR(p) model with CCC-GARCH noises for X_t and ARMA(p,q)-GARCH for y_t according to some information criterium, in order to replicate the case in which we do not count with any information about the DGP of X_t or y_t. We use the Bayesian Information Criterium (BIC). We also add an homoscedastic version of the selected model for y_t in order to capture the effect of disregarding the conditional heteroscedasticity in the prediction intervals.

For measuring accuracy in the prediction intervals, we generate $P = 1000$ trajectories of $X_{t+k|t}$, and obtain the empirical coverage of each model by counting the proportion of observations laying within each prediction interval. Thus, we obtain B coverages and calculate their mean absolute deviation (MAD) with respect to the nominal coverage. The model obtaining the smallest MAD will be the one yielding the most accurate prediction intervals, i.e. the one giving coverages closest to the nominal. We choose the forecast horizons $k = 1, 2, 6, 12, 24$ and a relatively small sample size $T = 200$ to resemble the case of a monthly series, such as the aggregate inflation rate. Table 5.2 shows the results for the 95% prediction intervals. The top block of rows correspond to the cases in which we know the DGP of the data, and the bottom block correspond to the case in which we need to find the best model and estimate it.

relationship, e.g. the sum of two dependent (strong) GARCH(1,1) processes, we just use the Monte Carlo averages of the QML estimates of the GARCH(1,1) fitted to the aggregate of very long simulated series. That is, we simulate 100 bivariate processes, X_t, with a sample size of $T = 20000$, find their aggregate by means of the vector λ, fit the GARCH(1,1) process and collect the ARCH and GARCH coefficients.

With respect to the results in which we know the DGP, the main conclusion is that the indirect approach of using the multivariate model provides more accurate prediction intervals than the direct approach of fitting the corresponding ARMA-GARCH model to the aggregate. This conclusion is maintained for all DGPs. However, as k increases, the volatility shocks vanish and consequently, the direct approach produce prediction intervals with coverages closer to those of the indirect approach.

With respect to the cases in which the DGP is unknown and the parameters need to be estimated, again we observe that in general the indirect approach yields the best coverages, for almost all DGPs. Of course, with respect to the homoscedastic models, the indirect approach greatly improves the coverages for small k. Again, as k increases, the indirect and the direct approaches with or without GARCH noises yield similar results. Among the different DGPs used to simulate the data, the ones with R being unrestricted and only the second component being GARCH (2^{nd}-F) yield the worst coverages in both, the indirect and direct approaches. In other words, the cases in which the homoscedastic noise is correlated with the heteroscedastic one produce relatively worse coverages. However, when we set the first component to be GARCH, both approaches yield much better coverages. This pattern may be explained by the fact that the approximated $\hat{H}_{t+k|t}$ given in (5.13) may not be good when one component is homoscedastic, although it may be valid if both noises are GARCH processes with the same parameters, as in our case.

Finally, we also carry out Monte Carlo experiments to measure the accuracy of the cumulative prediction intervals for all the DGPs considered in Table 5.1. Table 5.3 reports the results. Note that for $k = 1$ the results are identical to those of Table 5.2, by construction. In general the relative performance of the two approaches with respect to p_t is similar to the case of y_t for small k. However, as k increases, we see that the differences in accuracy of the different prediction intervals are larger. As mentioned before, this is a consequence of having volatility shocks that do not vanish with k. Thus, if the direct or indirect approach captures differently a given excess volatility, their conditional variances will differ for all k. Of course, given that the homoscedastic case does not take into account these shocks, it will always has the worst coverages for almost all DGPs.

As a general result of these Monte Carlo experiments, we can say that the univariate model for the aggregate does not generate as good prediction intervals as the indirect approach of using a multivariate model, mainly because the former is not able to distinguish the source of the volatility shocks. Consequently, it cannot properly assess the true uncertainty associated with the future values of the series. This is more evident when working with the cumulative forecasts.

5.5 Conclusions

In this chapter, we study the performance of the prediction intervals of the aggregate when at least one of its components is conditionally heteroscedastic. In particular, we analyze if using the indirect approach of modelling the whole vector of components and then aggregate the forecasts

improves the accuracy of prediction intervals when compared with the direct approach of taking the aggregate as a univariate series. We obtain the multi-step MSE point predictions and their conditional variance k steps ahead for the case of a VAR(p) process with a constant conditional correlation (CCC) GARCH residual vector. We find with simulated data that the use of the indirect approach may be preferred because it yields more accurate prediction intervals than the approach of working directly with the aggregate. Making an analogous argument as in Chapter 3, this result may be explained by the fact that the direct univariate approach is not able to distinguish the source of the volatility shocks, i.e. which components have excess volatilities different from zero. Consequently, it cannot properly assess the true uncertainty associated with the future values of the series.

	VAR(1) with CCC-GARCH(1,1) noises					ARMA(2,1) with GARCH(1,1) noises				
DGP Label	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$
U-B-D	0.518	0.565	0.560	0.589	0.609	1.689	1.558	1.262	0.974	0.731
F-B-D	0.518	0.579	0.574	0.628	0.604	2.035	1.956	1.487	1.111	0.788
D-B-D	0.547	0.580	0.613	0.616	0.602	2.006	1.868	1.595	1.280	0.903
U-2^{nd}-D	0.560	0.570	0.570	0.574	0.544	2.094	1.925	1.571	1.199	0.796
F-2^{nd}-D	0.535	0.527	0.562	0.574	0.549	2.300	2.266	1.866	1.416	0.940
D-2^{nd}-D	0.547	0.553	0.597	0.583	0.586	2.143	2.047	1.708	1.330	0.899
U-B-F	0.570	0.547	0.579	0.562	0.567	0.915	0.870	0.784	0.668	0.637
F-B-F	0.552	0.561	0.582	0.584	0.575	0.998	0.948	0.843	0.758	0.635
D-B-F	0.561	0.575	0.581	0.584	0.587	1.031	1.017	0.944	0.800	0.673
U-2^{nd}-F	0.538	0.572	0.565	0.573	0.557	2.372	2.164	1.589	1.112	0.732
F-2^{nd}-F	0.550	0.538	0.551	0.534	0.532	2.312	2.321	1.879	1.279	0.818
D-2^{nd}-F	0.567	0.545	0.595	0.562	0.564	2.272	2.074	1.583	1.245	0.826

	VAR(p)-MGARCH(1,1) (BIC)					ARMA(p,q)-GARCH(1,1) (BIC)					ARMA(p,q)-Homoscedastic (BIC)				
DGP Label	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$
U-B-D	1.822	1.849	2.098	2.316	2.572	2.425	2.345	2.351	2.436	2.613	3.993	3.778	3.313	2.874	2.627
F-B-D	1.819	1.820	2.031	2.244	2.521	2.787	3.073	2.411	2.455	2.641	4.226	4.315	3.278	2.780	2.522
D-B-D	1.774	1.889	2.450	2.824	3.155	2.638	2.541	2.872	3.044	3.199	4.217	3.935	3.759	3.422	3.141
U-2^{nd}-D	2.680	2.543	2.537	2.559	2.624	2.976	2.816	2.795	2.759	2.768	3.736	3.620	3.292	2.924	2.601
F-2^{nd}-D	2.749	3.060	3.032	2.940	2.958	3.222	3.897	3.207	3.071	3.112	3.638	3.931	3.132	2.876	2.828
D-2^{nd}-D	2.695	2.449	2.615	2.767	2.958	3.106	2.905	3.016	2.995	3.096	3.632	3.656	3.657	3.312	3.037
U-B-F	1.734	1.818	2.075	2.289	2.616	1.967	1.940	2.135	2.372	2.684	4.245	3.935	3.401	2.960	2.751
F-B-F	1.747	1.832	2.046	2.253	2.533	2.239	2.665	2.417	2.562	2.849	4.383	4.515	3.639	3.068	2.748
D-B-F	1.785	1.891	2.349	2.684	3.033	2.058	1.986	2.443	2.756	3.076	4.250	3.906	3.667	3.385	3.132
U-2^{nd}-F	3.172	3.093	2.915	2.920	2.997	3.289	3.108	3.037	3.078	3.182	3.752	3.667	3.434	3.215	3.045
F-2^{nd}-F	2.946	3.190	3.035	2.956	2.960	3.117	3.882	3.204	3.112	3.131	3.720	3.945	3.279	3.048	2.935
D-2^{nd}-F	3.058	2.873	2.853	3.059	3.251	3.328	2.990	3.167	3.271	3.418	3.728	3.637	3.752	3.528	3.290

Table 5.2: Mean absolute deviation (MAD) of the observed coverage, with respect to the nominal 95%. The data generating processes (DGPs) are described in Table 5.1 and the sample size selected is $T = 200$. The two models in the upper part correspond to the cases in which the DGP is known, while the three models in the bottom part correspond to the cases in which the research does not have information about the DGP and select the model according to the Bayesian information criterium (BIC).

	VAR(1) with CCC-GARCH(1,1) noises					ARMA(2,1) with GARCH(1,1) noises				
DGP Label	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$
U-B-D	0.518	0.571	0.603	0.606	0.598	1.689	1.610	1.506	1.385	1.165
F-B-D	0.518	0.5573	0.5798	0.6108	0.597	2.035	2.108	2.124	1.9074	1.5337
D-B-D	0.547	0.5788	0.6217	0.6194	0.5924	2.006	2.034	1.993	1.847	1.529
U-2^{nd}-D	0.560	0.583	0.579	0.586	0.591	2.094	1.994	1.808	1.618	1.352
F-2^{nd}-D	0.535	0.542	0.552	0.544	0.542	2.300	2.263	2.332	2.092	1.628
D-2^{nd}-D	0.547	0.550	0.602	0.598	0.627	2.143	2.106	2.061	1.889	1.565
U-B-F	0.570	0.586	0.581	0.578	0.572	0.915	0.874	0.865	0.841	0.786
F-B-F	0.552	0.608	0.588	0.593	0.551	0.998	1.039	1.157	1.157	0.998
D-B-F	0.561	0.597	0.606	0.605	0.612	1.031	1.032	1.145	1.185	1.017
U-2^{nd}-F	0.538	0.572	0.582	0.587	0.581	2.372	2.285	1.874	1.634	1.228
F-2^{nd}-F	0.550	0.545	0.541	0.578	0.539	2.312	2.281	2.354	2.072	1.574
D-2^{nd}-F	0.567	0.564	0.592	0.636	0.589	2.272	1.972	1.692	1.513	1.304

	VAR(p)-MGARCH(1,1) (BIC)					ARMA(p,q)-GARCH(1,1) (BIC)					ARMA(p,q)-Homoscedastic (BIC)				
DGP Label	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$	$k=1$	$k=2$	$k=6$	$k=12$	$k=24$
U-B-D	1.822	1.973	2.716	3.404	4.113	2.425	2.439	3.153	4.042	5.054	3.993	3.861	4.169	4.745	5.354
F-B-D	1.819	1.888	2.508	2.746	2.957	2.787	3.250	2.765	2.852	3.223	4.226	4.478	3.718	3.272	3.365
D-B-D	1.774	1.923	2.751	3.728	4.710	2.638	2.649	3.416	4.595	5.865	4.217	4.090	4.469	5.243	6.126
U-2^{nd}-D	2.680	2.668	3.163	3.717	4.299	2.976	2.903	3.501	4.295	5.167	3.736	3.693	4.182	4.765	5.376
F-2^{nd}-D	2.749	3.134	4.251	4.347	4.242	3.222	4.010	3.468	2.958	3.313	3.638	4.005	3.331	2.967	3.280
D-2^{nd}-D	2.695	2.510	2.828	3.591	4.482	3.106	3.051	3.613	4.580	5.628	3.632	3.661	4.200	4.923	5.737
U-B-F	1.734	1.911	2.743	3.526	4.312	1.967	1.988	2.725	3.612	4.672	4.245	3.971	4.153	4.559	5.182
F-B-F	1.747	1.888	2.641	2.987	3.260	2.239	2.803	2.451	2.695	3.087	4.383	4.679	3.972	3.226	3.156
D-B-F	1.785	1.973	2.779	3.623	4.549	2.058	2.039	2.901	4.293	5.875	4.250	4.021	4.313	5.224	6.338
U-2^{nd}-F	3.172	3.248	3.751	4.450	5.068	3.289	3.211	3.726	4.639	5.568	3.752	3.725	4.321	5.021	5.760
F-2^{nd}-F	2.946	3.251	4.048	4.108	4.082	3.117	3.992	3.236	2.844	3.270	3.720	4.028	3.190	2.902	3.233
D-2^{nd}-F	3.058	2.939	3.359	4.114	4.997	3.328	2.930	3.614	4.970	6.346	3.728	3.745	4.433	5.538	6.681

Table 5.3: Mean absolute deviation (MAD) of the observed coverage for the cumulative prediction intervals, with respect to the nominal 95%. See Table 5.2 for further information.

Chapter 6

Conclusions and future research

Stochastic trends and volatility clustering are two features that are present in many economic and financial time series. In this thesis, we study two alternative models to deal with these features, the unobserved component model with conditionally heteroscedastic noises and the ARIMA-GARCH model. We focus mainly on the local level model and its reduced form IMA counterpart and analyze the relative forecasting performance of the prediction intervals constructed from these two models. Throughout the chapters, we use either simulated or real time series data to illustrate the main results.

In Chapter 2 we derive the statistical properties of the reduced form unobserved component model when we allow the noises to be conditionally heteroscedastic and show that although working with the ARIMA reduced form model is simpler because there is only one disturbance, working with the unobserved component model may lead to discover conditionally heteroscedastic structures that could not be apparent in the reduced form noise.

When focusing of the relative forecasting performance of the two alternative approaches, we show in Chapter 3 that if the unobserved component models have conditionally heteroscedastic noises only in the transitory component, then the prediction intervals based on the homoscedastic and heteroscedastic unobserved components models stick to each other for large prediction horizons. However, depending on the sign of the excess volatility, the ARIMA-GARCH counterparts may be wider or thinner than the intervals obtained with the corresponding unobserved component model. This is due to its incapacity of distinguishing whether the heteroscedasticity affects the long or the short run components, and it may lead to significant differences between the two prediction intervals, specially for medium and long term. Therefore, the use of reduced form ARIMA models to construct prediction intervals may be inappropriate to capture the underlying uncertainty of the heteroscedastic components.

On the other hand, in Chapter 4 we compare the conditionally heteroscedastic local level model with a time-varying IMA model based on the rolling-window estimation approach. These models models have been recently used to forecast the dynamics of the US inflation. We show that the evolution of the MA parameter can only be attributed to breaks and not to the time-varying variances of the local level noises. Moreover, we can conclude that it is very difficult

to assess if the evolution of the rolling window estimator comes only from sample variability or from changes in the marginal variances of the local level noises, and that the rolling window estimation approach is not able to capture the conditionally heteroscedastic structure of the local level noises.

Finally, in Chapter 5 we study the forecasting performance of prediction intervals constructed from two alternative approaches of fitting an index series. In particular, we analyze if the use of a disaggregation approach consisting on fitting the components of the index series as a vector an then aggregate their predictions is better than the more direct approach of using a univariate model to fit and forecast the aggregate. We show with some simulated data that the latter does not generate as good prediction intervals as the former approach of using a multivariate model, mainly because the former is not able to distinguish the source of the volatility shocks. Consequently, it cannot properly assess the true uncertainty associated with the future values of the series.

Different topics have been arising while working on the different chapters of this thesis, which are part of the future research agenda. We list the most relevant ones below.

The study of the statistical properties of the conditionally heteroscedastic unobserved component models and their reduced form ARIMA counterpart given in Chapter 2 is made only for univariate time series. We believe that it may be worthy to generalize the results obtained for the case in which y_t is a vector, and compare the multivariate versions of the unobserved components with vector ARIMA processes, when some of the component noises are conditionally heteroscedastic. This may be of empirical interest specially when dealing with financial time series.

When working with simulated time series, we measure accuracy in prediction intervals by means of the observed coverage, which is computed as the percentage of simulated trajectories that fall inside the intervals. However, when dealing with real time series, the observed coverage must be estimated in an out-of-sample scheme. Consequently, to make inference about the relative accuracy of these intervals, we need to use some of the existing tests for interval and density forecasts; see, for example, Giacomini and Komunjer (2005), Corradi and Swanson (2005) and Giacomini and White (2006). Many of them are generalizations of the mean square forecast error (MSFE). We plan to include the results of these tests in the empirical illustrations.

In Chapter 4, we plan to make a more exhaustive empirical analysis of the forecasting performance of the presented models, by using other measures of the inflation rate, in order to check the robustness of the conclusions. On the other hand, in Chapter 5, we also wish to compare the two approaches using an index of the inflation rate, and decompose it into four of five components. Thus, we can identify which of them are conditionally heteroscedastic and then use the indirect approach of fitting the components as a vector time series, aggregate the predictions, and finally compare the relative accuracy of these predictions with those obtained directly from the univariate analysis of the aggregate inflation. Again, we should use some of the tests for interval and density forecasts mentioned above to perform the analysis.

Finally, as one of the main conclusions of the thesis is that using unobserved component

models to fit conditionally heteroscedastic time series may lead to improve accuracy on prediction intervals, we think that more effort should be made to find estimation procedures that yield estimators with better properties. In this regard, we think that the use of nonlinear filters or MCMC methods, some of them already existing in the literature, may be of great help.

Bibliography

Atkeson, A. and L. E. Ohanian (2001). Are phillips curves useful to forecasting inflation? *Federal Reserve Bank of Minneapolis Quarterly Review, 25*, 2–11.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 51*, 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics 72*, 498505.

Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). ARCH modelling in finance: A selective review of the theory and empirical evidence. *Journal of Econometrics 52*, 5–59.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. *The Handbook of Econometrics 2*, 2959–3038.

Bos, C., S. J. Koopman, and M. Ooms (2007). Long memory modelling of inflation with stochastic variance and structural breaks. Working Paper TI2007-099/4, Tinbergen Institute.

Bos, C. S. and N. Shephard (2006). Inference for adaptive time series models: Stochastic volatility and conditionally gaussian state space form. *Econometric Reviews 25*, 219–244.

Breidt, F. J. and R. A. Davis (1992). Time-reversibility, identifiability and independence of innovations for stationary time series. *Journal of Time Series Analysis 13*, 377 390.

Broto, C. and E. Ruiz (2006). Unobserved component models with asymmetric conditional variances. *Computational Statistics and Data Analysis 50*, 2146–2166.

Broto, C. and E. Ruiz (2009). Testing for conditional heteroscedasticity in the components of inflation. *Studies in Nonlinear Dynamics and Econometrics, forthcoming.*

Carnero, M. A., D. Peña, and E. Ruiz (2006). Effects of outliers on the identification and estimation of GARCH models. *Journal of Time Series Analysis, in press.*

Cecchetti, S. G., P. Hooper, B. C. Kasman, K. L. Schoenholtz, and M. W. Watson (2007). Understanding the evolving inflation process. Technical report, U.S. Monetary Policy Forum 2007 (February).

Chadha, J. S. and L. Sarno (2002). Short- and long-run price level uncertainty on investment under different monetary policy regimes: an international comparison. *Oxford Bulletin of Economics and Statistics 64*, 183–212.

Chang, K. H. and M. J. Kim (2004). Jumps and time-varying correlations in daily foreign exchange rates. *Journal of International Money and Finance 119*, 257–289.

Cogley, T. and T. J. Sargent (2007). Inflation-gap persistence in the u.s. Working paper, University of California, Davis.

Corradi, V. and N. R. Swanson (2005). Predictive density and conditional confidence interval accuracy tests. *Journal of Econometrics, 135*, 187–228.

Diebold, F. X. (2004). *Measuring and Forecasting Financial Market Volatilities and Correlations*. New York: Norton.

Diebold, F. X. and J. Lopez (1995). *Modelling Volatility Dynamics*, pp. 427–472. Boston : Kluwer Academic Press.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*, 253–263.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation on GARCH processes. *Econometrica 61*, 909–927.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford, UK: Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica 50*, 987–1008.

Espasa, A. and R. Albacete (2004). Econometric modelling for short-term inflation forecasting in the EMU. Working Paper 03-43, Universidad Carlos III.

Fisher, J., C. T. Liu, and R. Zhou (2002). When can we forecast inflation? *Federal Reserve Bank of Chicago Economic Perspectives 1Q/2002*, 30–42.

Giacomini, R. and I. Komunjer (2005). Evaluation and combination of conditional quantiles forecasts. *Journal of Business & Economic Statistics 23*, 416–431.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*, 15451578.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *The Review of Economic Studies 61*(2), 247–264.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C. and A. Jaeger (1993). Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics 8*, 231–247.

Harvey, A. C. and S. Koopman (1992). Diagnostic checking of unobserved-components time series models. *Journal of Business & Economic Statistics 10*, 377–389.

Harvey, A. C., E. Ruiz, and E. Sentana (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics 52*, 129–157.

He, C. and T. Terasvirta (2004). An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory 20*, 904–926.

Hendry, D. F. and K. Hubrich (2006). Forecasting economic aggregates by disaggregates. Working Paper 9628, European Central Bank.

Hlouskova, J., K. Schmidheiny, and M. Wagner (2009). Multistep predictions for multivariate garch models: Closed form solution and the value for portfolio management. *Journal of Empirical Finance 16*(2), 330 – 336.

Hubrich, K. (2005). Forecating euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? *International Journal of Forecasting 21*, 119–136.

King, M. A., E. Sentana, and E. Wadhwani (1994). Volatility and links between national stock markets. *Econometrica 62*, 901–933.

Koopman, S. J. and C. S. Bos (2004). State space models with a common stochastic variance. *Journal of Business & Economic Statistics 22*, 346–357.

Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard (2000). *STAMP: Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultants Press.

Li, W. K., S. Ling, and M. McAleer (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys 16*(3), 245–69.

Ling, S. and M. McAleer (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory 19*, 280–310.

Lütkepohl, H. (1984). Linear transformations of vector ARMA processes. *Journal of Econometrics 26*, 283–293.

Lütkepohl, H. (1987). *Forecasting aggregated vector ARMA processes*. Berlin: Springer-Verlang.

Maravall, A. (1983). An application of nonlinear time series forecasting. *Journal of Business & Economic Statistics 1*, 66–74.

McLeod, A. I. and W. K. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis 4*, 269–273.

Meddahi, N. and E. Renault (2004). Temporal aggregation of volatility models. *Journal of Econometrics 119*, 355–379.

Meyer, R. and J. Yu (2000). BUGS for a bayesian analysis of stock volatility models. *Econometrics Journal 3*, 198–215.

Moore, B. and L. Schaller (2002). Persistent and transitory shocks, learning and investment dynamics. *Journal of Money, Credit and Banking 34*, 650–677.

Müller, U. K. and M. W. Watson (2006). Testing models of low-frequency variability. Working Paper W12671, NBER.

Nelson, C. R. and G. W. Schwert (1977). Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest rate is constant. *American Economic Review 67*, 478–486.

Nijman, T. and E. Sentana (1996). Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics 71*, 71–87.

Nyblom, J. and A. Harvey (2001). Testing against smooth stochastic trends. *Journal of Applied Econometrics 3*, 415–429.

Orphanides, A. and S. van Norden (2005). The reliability of inflation forecast based on output gap estimates in real time. *Journal of Money, Credit and Banking 37*, 583–600.

Pascual, L., J. Romo, and E. Ruiz (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics and Data Analysis, 50*, 2293–2312.

Rodriguez, J. and E. Ruiz (2005). A powerful test for conditional heteroscedasticity for financial time series with highly persistent volatilities. *Statistica Sinica 15*, 505–525.

Sensier and van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *Review of Economics and Statistics 86*, 833–839.

Sentana, E. (2004). Factor representing portfolios in large asset markets. *Journal of Econometrics 119*, 257–289.

Sentana, E. and G. Fiorentini (2001). Identification, estimation and testing of conditionally heteroscedastic factor models. *Journal of Econometrics 102*, 143–164.

Silvestrini, A. and D. Veredas (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys 22*, 458–497.

Stock, J. H. and M. W. Watson (2002). *Has the business cycle changed, and why?* Cambridge, MIT Press.

Stock, J. H. and M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking 39*, 3.

Taylor, S. J. (1986). *Modelling Financial Time Series*. New York: John Wiley & Sons.

Zaffaroni, P. (2007). Contemporaneous aggregation of GARCH processes. *Journal of Time Series Analysis 28*, 521–544.